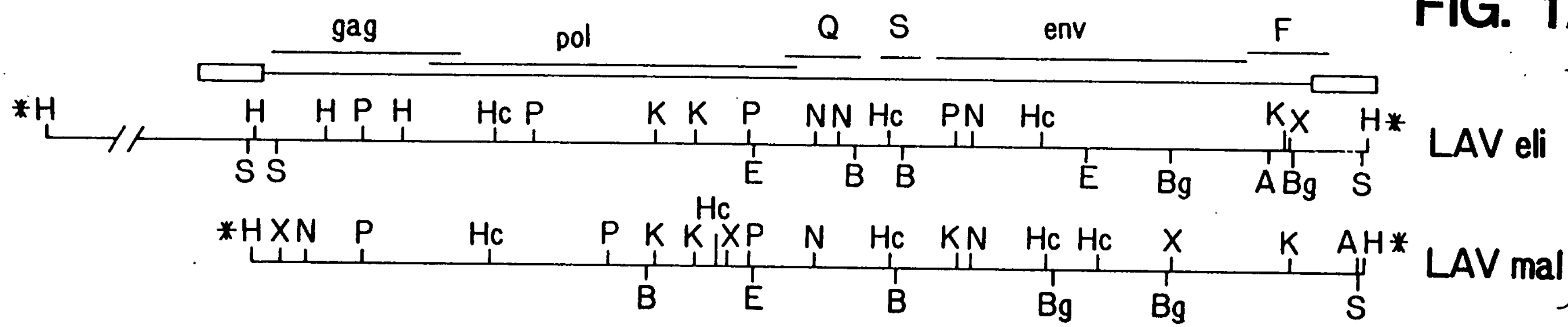
FIG. 1A


gag
pol
Q
S
env
F
*H
H H P H Hc P K K P NN Hc PN Hc KX H*
S S E B B E Bg A Bg S
LAV eli
H X N P Hc P K K Hc X P N Hc K N Hc Hc X K A H
B E B Bg Bg S
LAV mal

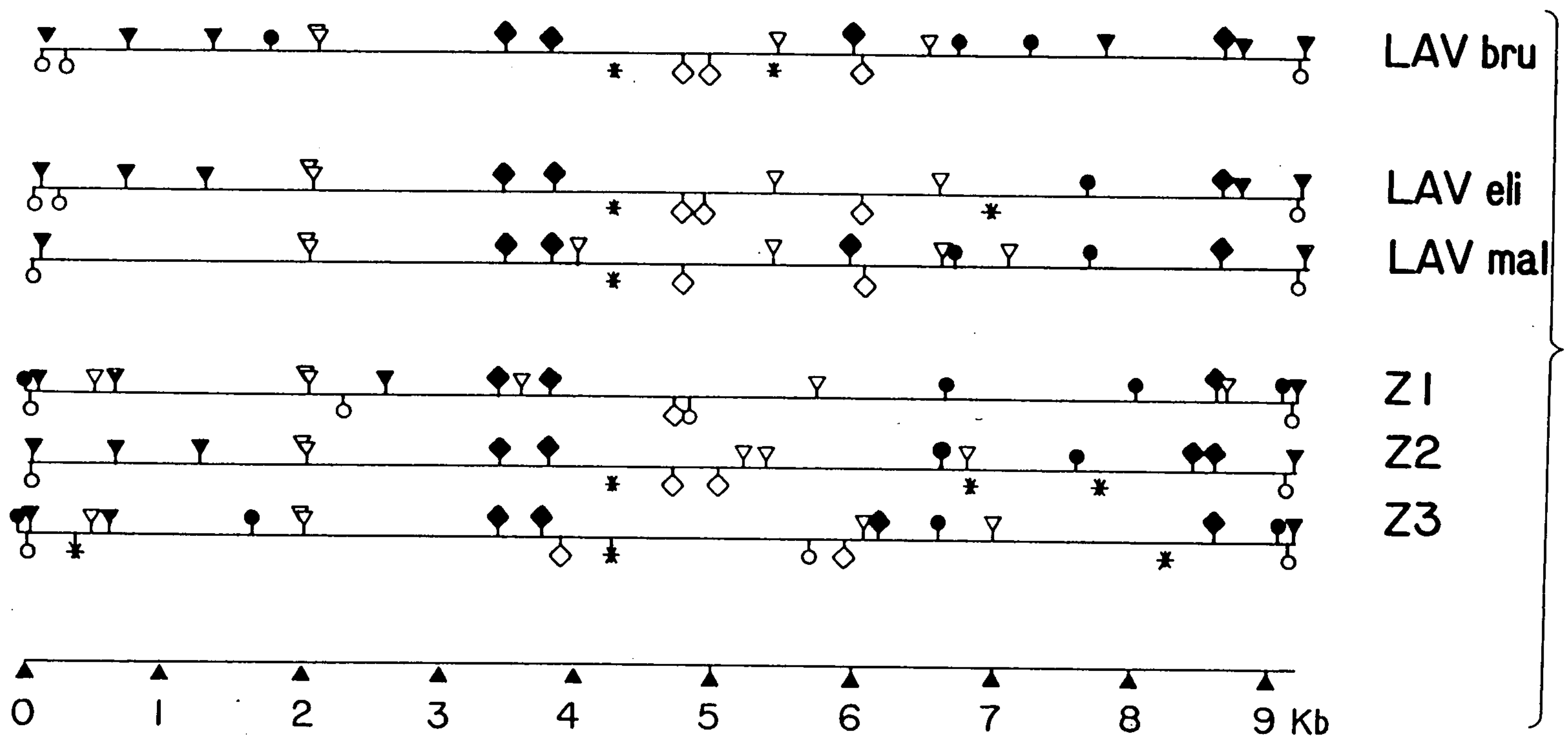
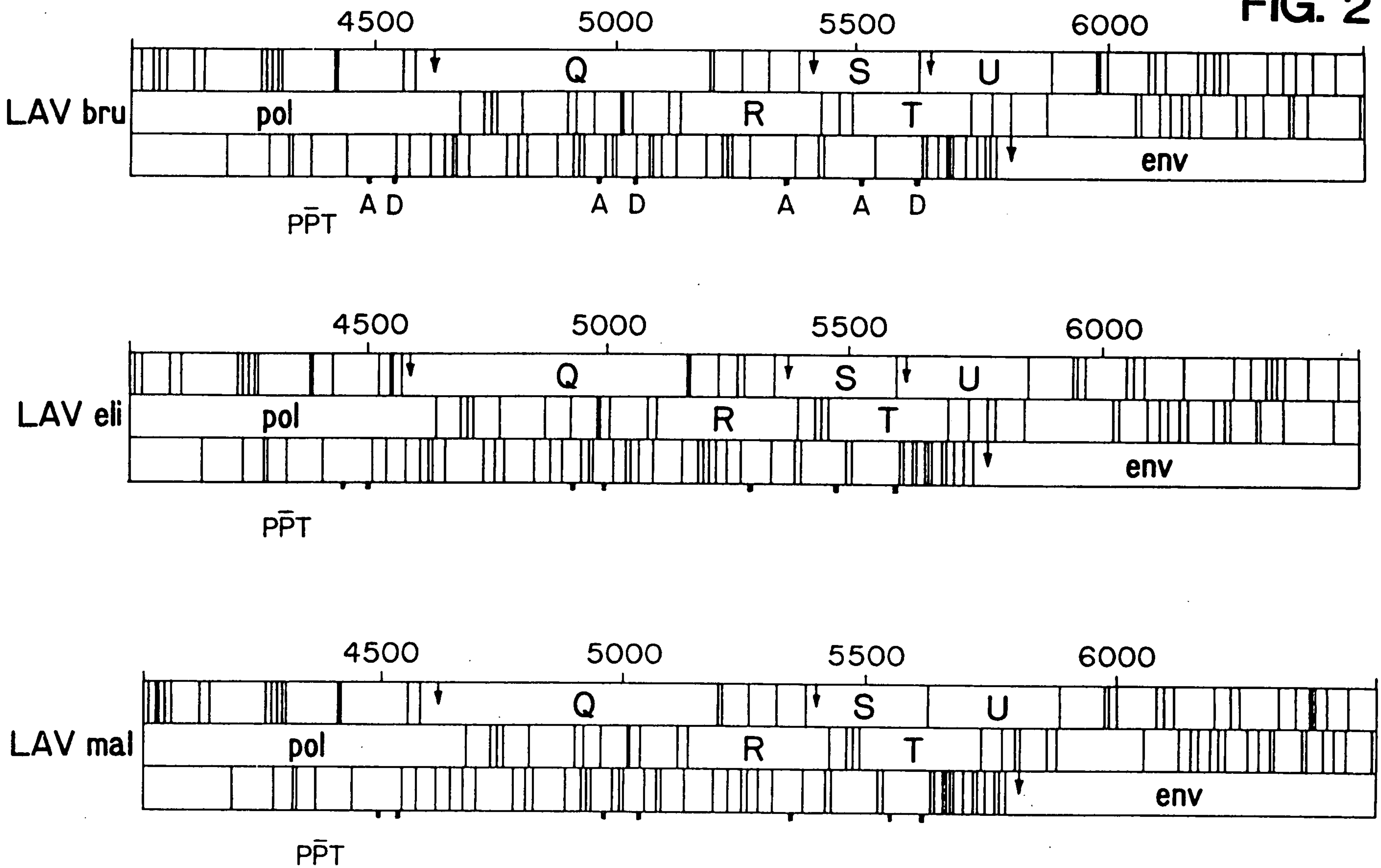


FIG. 1B

FIG. 2


4500
5000
5500
6000
LAV bru
Q
S
U
pol
R
T
env
A D
A D
A
A
D
PP̄T
LAV eli
PP̄T
LAV mal
PP̄T

GAG

```
                10         20         30         40         50         60         70         80
LAV BRU  MGARASVLSG GELDRWEKIR LRPGGKKKYK LKHIVWASRE LERFAVNPGL LETSEGCRQI LGQLQPSLQT GSEELRSLYN
ARV 2                   K
LAV MAL              K  A               R    L            L        C   Q   ME   ST K      IK
LAV ELI              K  K               R              Y L            K   I     AI    T

                                                                    p25
                                                                     ↓
                90        100        110        120        130        140        150        160
LAV BRU  TVATLYCVHQ RIEIKDTKEA LDKIEEEQNK SKKKAQQAAA ------DTGH SSQVSQNYPI VQNIQGQMVH QAISPRTLNA
ARV 2                 DV        E                    ----AAG  N               L
LAV MAL               DV             I     RQ T      AQQAAAA KN   S          A    I
LAV ELI           K G DV        E M                  ------   N N            L

               170        180        190        200        210        220        230        240
LAV BRU  WVKVVEEKAF SPEVIPMFSA LSCGATPQDL NTMLNTVGGH QAAMQMLKET INEEAAEWDK VHPVHAGPIA PGQMREPRGS
ARV 2
LAV MAL      I                             M   I             D        D             P
LAV ELI      I                                                           L

               250        260        270        280        290        300        310        320
LAV BRU  DIAGTTSTLQ EQIGWMTNNP PIPVGEIYKR WIILGLNKIV RMYSPTSILD IRQGPKEPFR DYVDRFYKTL RAEQASQEVK
ARV 2                                                                                       D
LAV MAL                    S        D                      V                               T
LAV ELI                A   S                 V             V                      F           D
```

FIG. 3A-1

```
                330        340        350        360        370        380     p13↓ 390        400
LAV BRU  NWMTETLLVQ NANPDCKTIL KALGPAATLE EMMTACQGVG GPGHKARVLA EAMSQVTNS- ATIMMQRGNF RNQRKIVKCF
ARV 2                                                                    P-   N               T
LAV MAL                             G                     S          A     T   A         KG - RI
LAV ELI                             Q                     S          A     V  T A        KGP   I

                410        420        430        440        450        460        470        480
LAV BRU  NCGKEGHIAR NCRAPRKKGC WKCGKEGHQM KDCTERQANF LGKIWPSYKG RPGNFLQSRP EPTAPPFLQS RPEPTAPPEE
ARV 2             K            R  R
LAV MAL          L                                         H                      ---- --------
LAV ELI           K            R       L              R    H                      ---- --------A
                                                                                  ---- --------A

                490        500        510
LAV BRU  SFRSGVETTT PSQKQEPIDK ELYPLTSLRS LFGNDPSSQ
ARV 2       F E K
LAV MAL    GF E IK-        QK       A  K         QL
LAV ELI    GF E I -        QK          K          L
```

FIG. 3A–2

Central Region: Q

```
                  10         20         30         40         50         60         70         80

LAV BRU   MENRWQVMIV WQVDRMRIRT WKSLVKHHMY VSGKARGWFY RHHYESPHPR ISSEVHIPLG DARLVITTYW GLHTGERDWH
ARV 2                                      I K  K           T    V            K               E
LAV MAL                          H           K  KN          R  K V               VR      Q   K
LAV ELI                      K               K  NR             K          E     K           E

                  90        100        110        120        130        140        150        160

LAV BRU   LGQGVSIEWR KKRYSTQVDP ELADQLIHLY YFDCFSDSAI RKALLGHIVS PRCEYQAGHN KVGSLQYLAL AALITPKKIK
ARV 2          A       K        G        H       E    KN I  YR                                 T
LAV MAL     H        Q      L   D                E     Q I          E                  T   A   TR
LAV ELI               R         G       M        E       I  D                          T   A  Q

                 170        180        190

LAV BRU   PPLPSVTKLT EDRWNKPQKT KGHRGSHTMN GH
ARV 2           K
LAV MAL         R            Q
LAV ELI         R            Q  R
```

FIG. 3B-1

```
                         R

                                10         20         30         40         50         60         70         80
LAV BRU    MEQAPEDQGP QREPHNEWTL ELLEELKNEA VRHFPRIWLH GLGQHIYETY GDTWAGVEAI IRILQQLLFI HFRIGCRHSR
ARV 2                     Y             R          P      S   Y                                  Q
LAV MAL          A                      Q                 S               E        S             Q
LAV ELI          A        Y    A        S                 S               V                      Q

                                90
LAV BRU    IGVTQQRRAR -NGASRS
ARV 2        II       R
LAV MAL      I R      -  S
LAV ELI      IIR      -  S

                S (tat)
                                10         20         30         40         50         60         70
LAV BRU    MEPVDPRLEP WKHPGSQPKT ACTTCYCKKC CFHCQVCFTT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQ
ARV 2           N             R     NN          YA   R  G                    A    D     A
LAV MAL      D   N      N     R   P NK       Y    M  I  G                         N A   DP P E
LAV ELI      D   N      N     R   P NK H     Y   P   LN G                   G     G A    PIP
```

FIG. 3B-2

```
POL

                  10         20         30         40         50         60         70         80
LAV BRU  FFREDLAFLQ GKAREFSSEQ TRANSPTFSS EQTRANSPTR RELQVWGRDN NSLSEAGADR QGTVSFNFPQ ITLWQRPLVT
ARV 2                                --- ----------         GE
LAV MAL      N   P       P           --- ----------S   R    G - KT   T  E   I    S           V
LAV ELI      N   P    G L PK         --- ----------S   R      -  P  KT  E                      A

                  90        100        110        120        130        140        150        160
LAV BRU  JKIGGQLKEA LLDTGADDTV LEEMSLPGRW KPKMIGGIGG FIKVRQYDQI LIEICGHKAI GTVLVGPTPV NIIGRNLLTQ
ARV 2     R                        N   K                      PV
LAV MAL  VRV                      IN   K                                K     I             M
LAV ELI                            N   K                      P         Q

                 170        180        190        200        210        220        230        240
LAV BRU  IGCTLNFPIS PIETVPVKLK PGMDGPKVKQ WPLTEEKIKA LVEICTEMEK EGKISKIGPE NPYNTPVFAI KKKDSTKWRK
ARV 2
LAV MAL                           R                T   KD       L
LAV ELI                                            T    D        R             I

                 250        260        270        280        290        300        310        320
LAV BRU  LVDFRELNKR TQDFWEVQLG IPHPAGLKKK KSVTVLDVGD AYFSVPLDED FRKYTAFTIP SINNETPGIR YQYNVLPQGW
ARV 2                                                     K
LAV MAL    N
LAV ELI                                                              S
```

FIG. 3C-1

```
                330        340        350        360        370        380        390        400
LAV BRU  KGSPAIFQSS MTKILEPFRK QNPDIVIYQY MDDLYVGSDL EIGQHRTKIE ELRQHLLRWG LTTPDKKHQK EPPFLWMGYE
ARV 2                                                                       F
LAV MAL                      T K  E                                  E   K  F
LAV ELI                           EM                             K   E      F  R

                410        420        430        440        450        460        470        480
LAV BRU  LHPDKWTVQP IVLPEKDSWT VNDIQKLVGK LNWASQIYPG IKVRQLCKLL RGTKALTEVI PLTEEAELEL AENREILKEP
ARV 2                M                            A     K
LAV MAL              Q  D E                             K          A   DIV     A
LAV ELI           S  K    E         N   ER

                490        500        510        520        530        540        550        560
LAV BRU  VHGVYYDPSK DLIAEIQKQG QGQWTYQIYQ EPFKNLKTGK YARTRGAHTN DVKQLTEAVQ KITTESIVIW GKTPKFKLPI
ARV 2      E          V                                  M                   VS           I
LAV MAL                                     QY           IKS                   AQ             R
LAV ELI                         H                        M             A     R S         R    R
```

FIG. 3C–2

```
                 570        580        590        600        610        620        630        640
LAV BRU   QKETWETWWT EYWQATWIPE WEFVNTPPLV KLWYQLEKEP IVGAETFYVD GAASRETKLG KAGYVTNRGR QKVVTLTDTT
ARV 2           A  M                                                N             D        SIA
LAV MAL         A                                 T                 N    K        D        S  E
LAV ELI            A                                   I            N             D        P

                 650        660        670        680        690        700        710        720
LAV BRU   NQKTELQAIH LALQDSGLEV NIVTDSQYAL GIIQAQPDKS ESELVNQIIE QIIKKEKVYL AWVPAHKGIG GNEQVDKLVS
ARV 2                                                      S
LAV MAL                      S                          I          Q D     S
LAV ELI            N

                 730        740        750        760        770        780        790        800
LAV BRU   AGIRKVLFLD GIDKAQDEHE KYHSNWRAMA SDFNLPPVVA KEIVASCDKC QLKGEAMHGQ VDCSPGIWQL DCTHLEGKVI
ARV 2              N       E                                                                 I
LAV MAL   S                E                     I                                           I
LAV ELI   Q                E          N

                 810        820        830        840        850        860        870        880
LAV BRU   LVAVHVASGY IEAEVIPAET GQETAYFLLK LAGRWPVKTI HTDNGSNFTS TTVKAACWWA GIKQEFGIPY NPQSQGVVES
ARV 2
LAV MAL   I                            I           VV            AA         N
LAV ELI                                            VV            AA
```

FIG. 3D-1

```
                 890        900        910        920        930        940        950        960
LAV BRU   MNKELKKIIG QVRDQAEHLK TAVQMAVFIH NFKRKGGIGG YSAGERIVDI IATDIQTKEL QKQITKIQNF RVYYRDSRDP
ARV 2       N                                                                                KK
LAV MAL                 E                                    I M                             N
LAV ELI                                    RR            I                  I

                 970        980        990       1000       1010
LAV BRU   LWKGPAKLLW KGEGAVVIQD NSDIKVVPRR KAKIIRDYGK QMAGDDCVAS RQDED
ARV 2
LAV MAL   I                                                    G G
LAV ELI   I                     K          V
```

FIG. 3D-2

ENV

```
                                           SP                                                    OMP

                  10         20         30         40         50         60         70         80
LAV BRU   MRVK---EKY QHLWRWGWKW GTMLLGILMI CSATEKLWVT VYYGVPVWKE ATTTLFCASD AKAYDTEVHN VWATHACVPT
ARV 2      K  GTRRN         --- - L   M                                      R
LAV MAL      REIQRN   NW    --- -M    M T    IA D                             S E      I
LAV ELI     ARGIERNC  NW K  --- -I      T     ADN                             S E  A   I

                  90        100        110        120        130        140        150        160
LAV BRU   DPNPQEVVLV NVTENFNMWK NDMVEQMHED IISLWDQSLK PCVKLTPLCV SLKCTDL-CN ATNTNSSNTN SSSGEMMME-
ARV 2              C             N     Q                         T N    - K        --- ----NWKE I
LAV MAL         IE E     G       N                               T N  NVN T  V GTNACS  RTNA LK  I
LAV ELI         IA E             N                               T N S E--L RN GTMG NV TTEEKG----

                 170        180        190        200        210        220        230        240
LAV BRU   KGEIKNCSFN ISTSIRGKVQ KEYAFFYKLD IIPIDNDTTS -----YTLTS CNTSVITQAC PKVSFEPIPI HYCAPAGFAI
ARV 2                 T    D I    N L RN   VV   AS  T TNYTN R IN   R                      T
LAV MAL   -  V        TPVGSD R  -   T  N   LVQ  DSDN  ----S R IN               T D
LAV ELI   ---M       VT VLKD K  QV  L  R    V     SST -NSTN R IN     A

                 250        260        270        280        290        300        310        320
LAV BRU   LKCNKKTFNG TGPCTNVSTV QCTHGIRPVV STQLLLNGSL AEEEVVIRSA NFTDNAKTII VQLNQSVEIN CTRPNNNTRK
ARV 2                K                  I                      D    N           E  A
LAV MAL       D K     EI K            K                   IM   E  L   T N       ET T       G    R
LAV ELI      RD K                                          I   E  L N   N   AH  E  K T  A  YQ   Q
```

FIG. 3E–1

```
                330        340        350        360        370        380        390        400
                                        •                                                      •
LAV BRU  SIRIQRGPGR AFVTIGK-IG NMRQAHCNIS RAKWNATLKQ IASKLREQFG NNKT-IIFKQ SSGGDPEIVT HSFNCGGEFF
ARV 2       Y --        W T RI  DI K         Q  N  E    VK                                  
LAV MAL  G HF--    Q  LY T I-V  DI R Y T N ETE DK  Q   V V  GSLL- -   - V N          M      R
LAV ELI  RTP -- L  Q SLY TKS-RS IIG           Q SK  Q   V R  GTLL- -   K   NS        T      R
                                                                       I K  P        T

                410        420        430        440        450        460        470        480
                                                   •                                •
LAV BRU  YCNSTQLFNS TWFNSTWSTE CSNNTEGSDT ITLPCRIKQF INMWQEVGKA MYAPPISGQI RCSSNITGLL LTRDGGNN--
ARV 2        T    N     -----RLN RTEG K N    I       I                    C  S                T -V
LAV MAL      TSK        Q NGARL- -  S  STGS          I       KT            A V  N L        I   NSSD
LAV ELI      TSG        NI A NNI TES NSTNTN     Q     I  K VAGR-     I     ERN  L              I --

                490        500        510        520        530↓       540   TMP  550        560
LAV BRU  NNGSEIFRPG GGDMRDNWRS ELYKYKVVKI EPLGVAPTKA KRRVVQREKR AVGI-GALFL GFLGAAGSTM GARSMTLTVQ
ARV 2    T DT V                      I       I                        V  M                      
LAV MAL   SDN TL              I      R                     E      I L-  M             V L
LAV ELI   STN T                      Q            R        E      I L-  M             A L
                                                                                        V
```

FIG. 3E–2

09041925.031398

```
                  570        580        590        600        610        620        630        640
                                                                          •     •
LAV BRU    ARQLLSGIVQ QQNNLLRAIE AQQHLLQLTV WGIKQLQARI LAVERYLKDQ QLLGIWGCSG KLICTTAVPW NASWSNKSLE
ARV 2                                             W          R
LAV MAL                                           W          Q    R   M       H   F    S    R  D
LAV ELI       M                                                               H   N    S    R  N

                  650        660        670        680        690        700        710        720
LAV BRU    QIWNNMTWME WDREINNYTS LINSLIEESQ NQQEKNEQEL LELDKWASLW NWFNITNWLW YIKIFIMIVG GLVGLRIVFA
ARV 2      D  D     Q E    D   N T YT L                            S
LAV MAL    D        Q EK   S   G I YN        I     K                S SK     R   IV    I    I
LAV ELI    E  Q       E    D   G  Y         T     K                S  Q          I    I

                  730        740        750        760        770        780        790        800
                                                                                   •
LAV BRU    VLSIVNRVRQ GYSPLSFQTH LPTPRGP-DR PEGIEEEGGE RDRDRSIRLV NGSLALIWDD LRSLCLFSYH RLRDLLLIVT
ARV 2                          R   V   -    D                 V    D F     E         R         
LAV MAL       L             L  L       P             QG G           FS         N             AA
LAV ELI       L                L   A   -       T      G       V  L  FS                        A
                                                                                         I  AV

                  810        820        830        840        850        860        870
LAV BRU    RIVELLGRRG WEALKYWWNL LQYWSQELKN SAVSLLNATA IAVAEGTDRV IEVVQGA[C]RA IRHIPRRIRQ GLERILL
ARV 2       T  I  K           S      I           W        T           A R Y   L  H          L
LAV MAL                     L      G      I   T          [C]       IG RFG   L          F  A
LAV ELI                DI   L         R    S  FD I                   II R    VLN                S
```

FIG. 3F-1

F

```
                  10         20         30         40         50         60         70         80
LAV BRU   MGGKWSKSSV VGWPTVRERM R----RAEPA ADGVGAASR- ----DLEKUG AITSSNTAAT NAACAWLEAQ EE-EEVGFPV
ARV 2            R M G  SAI      RAEP            V  - ----                    D          -
LAV MAL            I     KI   I  ---- TP T ET    V QD AVSQ  D C   AA  SP  N   S --- PP   E
LAV ELI            I     AI   I  ---- TM         V  - ----               S    D          SD

                  90        100        110        120        130        140        150        160
LAV BRU   TPQVPLRRHT YKAAVDLSHF LKEKGGLEGL IHSQRRQDIL DLWIYUTQGY FPDWQNYTPC PGVRYPLTFG WCYKLVPVEP
ARV 2     R              L  I               W     E                           I          F
LAV MAL   R            G F             D   VW PK  E      V                    I F        F    HS
LAV ELI   R            E L                  W KK  E      V N   I              I           E    D

                 170        180        190        200        210
LAV BRU   DKVEEANKGE NTSLLHPVSL HGMDDPEREV LEWRFDSRLA FHHVARELHP EYFKNC
ARV 2     E      E    N     M      E A K    V     K      M         Y D
LAV MAL   EE     E    NC    I Q    E A      K K   S   LR R   Q     Y D
LAV ELI   QE   DTE   TN     ICQ    E    Q   K   N      E K   M    FY  -
```

FIG. 3F-2

FIG. 4A

A LAVbru vs.

LAVbru vs.	GAG		POL		ENV TOTAL		ENV OMP		ENV TMP	
HTLV-3 USA	512 0/0	0.8	1015 0/0	1.3	856 5/0	1.4	507 5/0	1.6	349 0/0	1.1
ARV-2 USA	502 12/2	3.4	1003 12/0	3.1	855 17/11	13.0	505 17/10	14.3	350 0/1	11.2
LAVeli ZAIRE	500 13/1	9.8	1002 13/0	5.5	853 22/14	20.7	504 22/14	25.3	349 0/0	13.8
LAVmal ZAIRE	505 14/7	12.0	1002 13/0	7.7	859 13/11	21.7	509 13/10	26.4	350 0/1	14.9

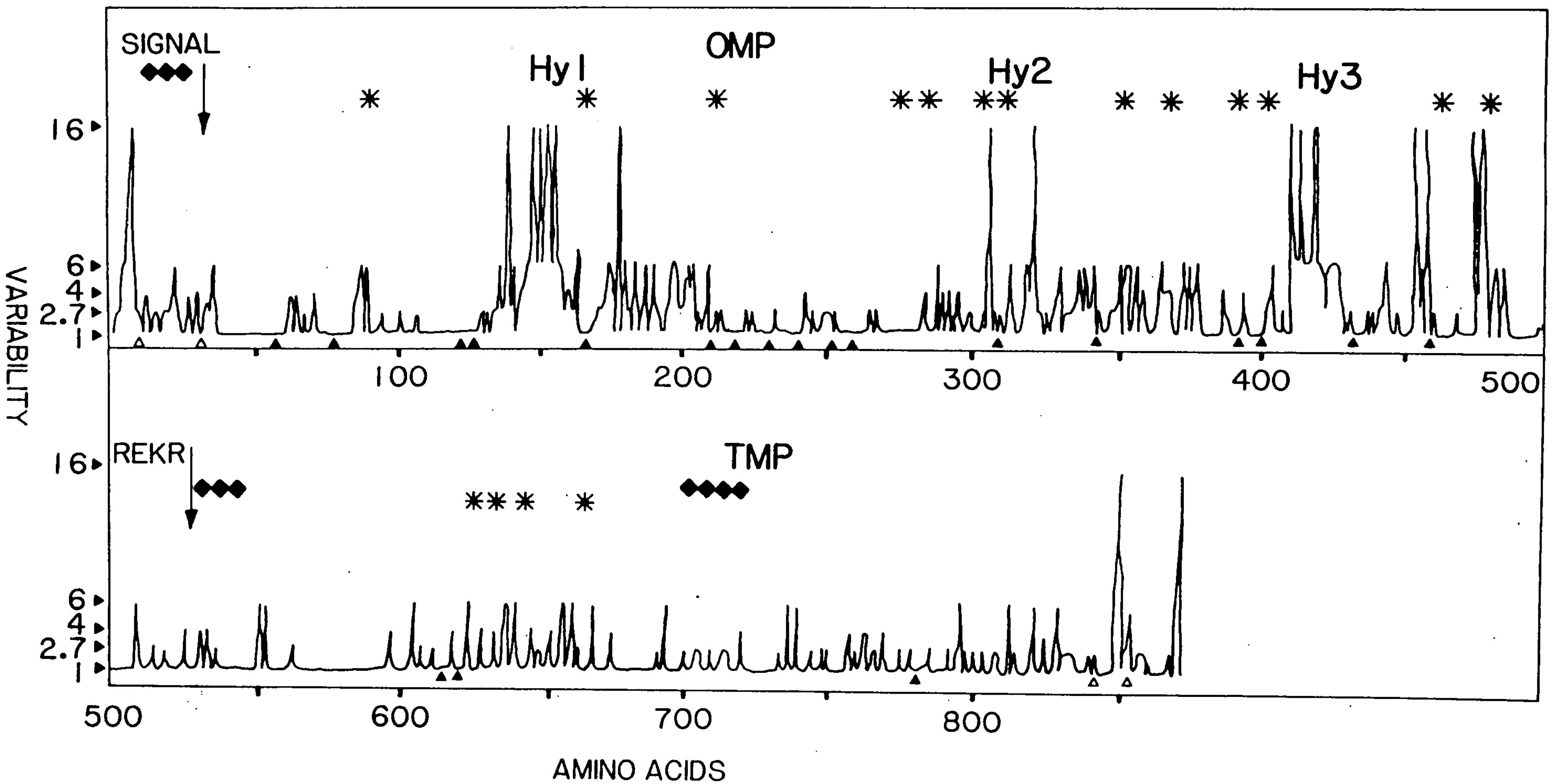
B LAVeli vs.

LAVeli vs.	GAG		POL		ENV TOTAL		ENV OMP		ENV TMP	
LAVmal	505 1/6	10.8	1002 0/0	8.4	859 13/11	19.8	509 8/13	23.6	350 0/1	14.3

FIG. 4B

A LAVbru vs.	orf F		central region					
			orf Q		orf R		orf S	
HTLV-3 USA	206 0/0	1.5	192 0/0	0		nd	80 0/0	2.5
ARV-2 USA	210 0/4	12.6	192 0/0	10.0	97 0/1	9.4	81 0/1	15.0
LAVeli ZAIRE	206 1/1	19.4	192 0/0	10.4	96 0/0	11.5	80 0/0	27.5
LAVmal ZAIRE	209 2/5	27.0	192 0/0	12.6	96 0/0	10.4	80 0/0	23.8
B LAVeli vs.								
LAVmal	209 3/6	22.5	192 0/0	12.0	96 0/0	6.3	80 0/0	11.3

FIG. 5


SIGNAL
Hy1
OMP
Hy2
Hy3
REKR
TMP
16
6
4
2.7
1
VARIABILITY
100
200
300
400
500
600
700
800
AMINO ACIDS

GAG

a

					120												
	K	A	Q	Q	A	A	A									D	T
LAV.BRU	AAA	GCA	CAG	CAA	GCA	GCA	GCT	-	-	-	-	-				GAC	ACA
	K	A	Q	Q	A	A	A	A	A							G	T
ARV 2	AAG	GCA	CAG	CAA	GCA	GCA	GCT	GCA	GCT	-	-	-	-			GGC	ACA
	K	T	Q	Q	A	A	A	A	Q	Q	A	A	A	A			T
LAV.MAL	AAG	ACA	CAG	CAG	GCA	GCA	GCT	GCA	CAG	CAG	GCA	GCA	GCT	GCC			ACA
	X	A	Q	Q	A	A	A									D	T
LAV.ELI	AAG	GCA	CAG	CAA	GCA	GCA	GCT	-	-	-	-	-	-			GAC	ACA

FIG. 6A–1

b

						460												470									480
LAV.BRU																											
G	N	F	L	Q	S	R	P	E	P	T	A	P	P	F	L	Q	S	R	P	E	P	T	A	P	P	E	E
GGG	AAT	TTT	CTT	CAG	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	TTT	CTT	CAG	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	GAA	GAG
ARV 2																											
G	N	F	L	Q	S	R	P	E	P	T	A	P	P													E	E
GGG	AAT	TTT	CTT	CAG	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	-	-	-	-	-	-	-	-	-	-	-	-	GAA	GAG
LAV.MAL																											
G	N	F	L	Q	S	R	P	E	P	T	A	P	P													A	E
GGG	AAT	TTC	CTT	CAG	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	-	-	-	-	-	-	-	-	-	-	-	-	GCA	GAG
LAV.ELI																											
G	N	F	L	Q	S	R	P	E	P	T	A	P	P													A	E
GGG	AAC	TTT	CTC	CAA	AGC	AGA	CCA	GAG	CCA	ACA	GCC	CCA	CCA	-	-	-	-	-	-	-	-	-	-	-	-	GCA	GAG

FIG. 6A–2

c

		20									30
	R	M	R					R	A	E	P	A
LAV.BRU	AGA	ATG	AGA	-	-	-	-	CGA	GCT	GAG	CCA	GCA
	R	M	R	R	A	E	P	R	A	E	P	A
ARV 2	AGA	ATG	AGA	CGA	GCT	GAG	CCA	CGA	GCT	GAG	CCA	GCA
	R	I	R					R	T	P	P	T
LAV.MAL	AGA	ATA	AGA	-	-	-	-	CGA	ACT	CCC	CCA	ACA
	R	I	R					R	T	P	P	T
LAV.ELI	AGA	ATA	AGA	-	-	-	-	AGA	ACT	AAT	CCA	GCA

d

					40							
	V	G	A	A	S	R						D
LAV.BRU	GTG	GGA	GCA	GCA	TCT	CGA	-	-	-	-	-	GAC
	V	G	A	V	A	R						D
ARV 2	GTG	GGA	GCA	GTA	TCT	CGA	-	-	-	-	-	GAC
	V	G	A	V	S	R	D	A	V	S	Q	D
LAV.MAL	GTA	GGA	GCA	GTA	TCT	CAA	GAT	GCA	GTA	TCT	CAA	GAT
	V	G	A	V	S	R						D
LAV.ELI	GTA	GGA	GCA	GTA	TCT	CGA	-	-	-	-	-	GAC

FIG. 6A-3

ENV

e

```
                                                    20
          Q   H   L   W   R   W   G   W   K   W   G   T   M   L
LAV.BRU  CAG CAC TTG TGG ACA TGG GGC TGG AAA TGG GGC ACC ATG CTC

          Q   H   L   W   R   W   G                   T   L   L
ARV 2    CAG CAC TTG TGG AGA TGG GGC  -   -   -   -  ACC TTG CTC

          Q   N   W   W   R   W   G                   M   M   L
LAV.MAL  CAA AAC TGG TGG AGA TGG GGC  -   -   -   -  ATG ATG CTC

          Q   N   W   W   K   W   G                   T   M   L
LAV.ELI  CAA AAC TGG TGG AAA TCG GGC  -   -   -   -  ATC ATG CTC
```

FIG. 6B-1

f

```
LAV.BRU                        140                                        150
 L   K   C   T   D   L       G   N   A   T   N   T   N   S   S   N   T   N   S   S   S   G   E
TTA AAG TGC ACT GAT TTG  -  GGG AAT GCT ACT AAT ACC AAT AGT AGT AAT ACC AAT AGT AGT AGC GGG GAA
 M   M   M   E       K   G   E   I
ATG ATG ATG GAG  -  AAA GCA GAG ATA

ARV 2
 L   N   C   T   D   L       G   K   A   T   N   T   N   S   S                               M
TTA AAT TGC ACT GAT TTG  -  GGG AAG GCT ACT AAT ACC AAT AGT AGT  -   -   -   -   -   -   -  AAT
 W   K   E   E   I   K   G   E   I
TGG AAA GAA GAA ATA AAA GGA GAA ATA
```

LAV.MAL

L	N	C	T	N	V	N	G	T	A	V	N	G	T	N	A	G	S	N	R	T	N	A	E
TTA	AAC	TGC	ACT	AAT	GTG	AAT	GGG	ACT	GCT	GTG	AAT	GGG	ACT	AAT	GCT	GGG	AGT	AAT	AGG	ACT	AAT	GCA	GAA
L	K	M	E	I		G	E	V															
TTG	AAA	ATG	GAA	ATT	-	GGA	GAA	GTG															

LAV.ELI

L	N	C	S	D	E			L	R	N	N	G	T	M	G	N	N	V	T	T	E	E	K
TTA	AAC	TGT	AGT	GAT	GAA	-	-	TTG	AGG	AAC	AAT	GGC	ACT	ATG	GGG	AAC	AAT	GTC	ACT	ACA	GAG	GAG	AAA
G								M															
GGA	-	-	-	-	-	-	-	ATG															

FIG. 6B-2

```
g                                  200
             D   N   D   T   T   S                       Y   T   L
LAV.BRU     GAT AAT GAT ACT ACC AGC -   -   -   -   -   TAT ACG TTG

             D   N   A   S   T   T  |T   N   Y | T   N   Y | R   L
ARV 2       GAT AAT GCT AGT ACT ACT |ACC AAC TAT|ACC AAC TAT| AGG TTG

            |D   D   S | D   N   S | S                       Y   R   L
LAV.MAL     |GAT GAT AGT|GAT AAT AGT| AGT -   -   -   -   -   TAT AGG CTA

             D   N   D   S  |S   T       N | S   T   N | Y   R   L
LAV.ELI     GAC AAT GAT AGT |AGT ACC  -  AAT|AGT ACC AAT| TAT AGG TTA
```

FIG. 6B-3

```
h
LAV.BRU                  410                                        420                 430
 C   N   S   T   Q   L  |F   N   S   T   W | F   N   S   T   W | S   T   E   G   S   N   N   T   E   G
TGT AAT TCA ACA CAA CTG |TTT AAT AGT ACT TGG|TTT AAT AGT ACT TGG| AGT ACT GAA GGG TCA AAT AAC ACT GAA GGA
 S   D   T   I
AGT GAC ACA ATC

ARV 2
 C   N   T   T   Q   L   F   N   N   T   W                       R   L   N   H  |T   K   G | T   K   G |
TGT AAT ACA ACA CAA CTG TTT AAT AAT ACA TGG -   -   -   -   -   AGG TTA AAT CAC |ACT GAA GGA|ACT AAA GGA|
 N   D   T   I
AAT GAC ACA ATC
```

```
LAV.MAL

 C   N   T   S   K   L   F   N   S   T   W   Q   N   N   G   A   R   L           S   N   S   T   E   S
TGT AAT ACA TCA AAA CTG TTT AAT AGT ACA TGG CAG AAT AAT GGT GCA AGA CTA  -   -  AGT AAT AGC ACA GAG TCA
 T   G   S   I
ACT GGT AGT ATC

LAV.ELI

 C   N   T   S   G   L   F  | N   S   T   W   N  | I   S   A   W   N  | N   I   T   E   S   N   N   S   T
TGT AAT ACA TCA GGA CTG TTT |AAT AGT ACA TGG AAT|ATT AGT GCA TGG AAT| AAT ATT ACA GAG TCA AAT AAT AGC ACA
 N   T   N   I
AAC ACA AAC ATC
```

FIG. 6B–4

LAV.MAL

→R
GGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGCAAGGAACCCACTG

R←→U5
CTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAAGCAGTGTGTGCCCATCTGTTGTGT

100 U5←
GACTCTGGTAACTAGAGATCCCTCAGACCACTCTAGACGGTGTAAAAATCTCTAGCAGTG

GCGCCCGAACAGGGACTTAAAGTGAAAGTAACAGGGACTCGAAAGCGGAAGTTCCAGAG

200
AAGTTCTCTCGACGCAGGACTCGGCTTGCTGAGGTGCACACAGCAAGAGGCGAGAGCGGC

→GAG 300
MetGlyAlaArg
GACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAG

AlaSerValLeuSerGlyGlyLysLeuAspAlaTrpGluLysIleArgLeuArgProGly
AGCGTCAGTATTAAGCGGGGGAAAATTAGATGCATGGGAGAAAATTCGGTTAAGGCCAGG

400
GlyLysLysLysTyrArgLeuLysHisLeuValTrpAlaSerArgGluLeuGluArgPhe
GGGAAAGAAAAAATATAGACTGAAACATTTAGTATGGGCAAGCAGGGAGCTGGAAAGATT

AlaLeuAsnProGlyLeuLeuGluThrGlyGluGlyCysGlnGlnIleMetGluGlnLeu
CGCACTTAACCCTGGCCTTTTAGAAACAGGAGAAGGATGTCAACAAATAATGGAACAGCT

500
GlnSerThrLeuLysThrGlySerGluGluIleLysSerLeuTyrAsnThrValAlaThr
ACAATCAACTCTCAAGACAGGATCAGAAGAAATTAAATCATTATATAATACAGTAGCAAC

600
LeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAlaLeuAspLysIle
CCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACCAAGGAAGCGCTAGATAAAAT

GluGluIleGlnAsnLysSerArgGlnLysThrGlnGlnAlaAlaAlaAlaGlnGlnAla
AGAGGAAATACAAAATAAGAGCAGGCAAAAGACACAGCAGGCAGCAGCTGCACAGCAGGC

700
AlaAlaAlaThrLysAsnSerSerSerValSerGlnAsnTyrProIleValGlnAsnAla
AGCAGCTGCCACAAAAAACAGCAGCAGTGTCAGTCAAAATTACCCCATAGTGCAAAATGC

GlnGlyGlnMetIleHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysVal
ACAAGGGCAAATGATACATCAGGCCATATCACCTAGGACTTTGAATGCATGGGTGAAAGT

800
IleGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGly
AATAGAAGAAAAGGCTTTCAGCCCAGAAGTGATACCCATGTTCTCAGCATTATCAGAGGG

900
AlaThrProGlnAspLeuAsnMetMetLeuAsnIleValGlyGlyHisGlnAlaAlaMet
GGCCACCCCACAAGATTTAAATATGATGCTGAACATAGTTGGAGGACACCAGGCAGCTAT

GlnMetLeuLysAspThrIleAsnGluGluAlaAlaAspTrpAspArgValHisProVal
GCAAATGTTAAAAGATACCATCAATGAGGAAGCTGCAGACTGGGACAGGGTACATCCAGT

1000
HisAlaGlyProIleProProGlyGlnMetArgGluProArgGlySerAspIleAlaGly
ACATGCAGGGCCTATTCCCCCAGGCCAGATGAGAGAACCAAGAGGAAGTGACATAGCAGG

FIG. 7A

```
 ThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrSerAsnProProIleProVal
AACTACTAGTACCCTTCAAGAACAAATAGGATGGATGACAAGCAACCCACCTATCCCAGT
                 1100
 GlyAspIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSer
GGGAGACATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAG
                                                        1200
ProValSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAsp
CCCTGTCAGCATTTTGGACATAAGACAAGGGCCAAAGGAACCTTTTAGAGACTATGTAGA

 ArgPhePheLysThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThr
TAGGTTCTTTAAAACTCTCAGAGCTGAGCAAGCTACACAGGAGGTAAAAAATTGGATGAC
                                     1300
 GluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGly
AGAAACCTTGCTGGTCCAAAATGCGAATCCAGACTGTAAGACCATTTTAAAAGCATTAGG

 ProGlyAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHis
ACCAGGGGCTACATTAGAAGAAATGATGACAGCATGCCAGGGAGTGGGAGGACCCAGTCA
                 1400
 LysAlaArgValLeuAlaGluAlaMetSerGlnAlaThrAsnSerThrAlaAlaIleMet
TAAAGCAAGAGTTTTGGCTGAGGCAATGAGCCAAGCAACAAATTCAACTGCTGCCATAAT
                                                        1500
 MetGlnArgGlyAsnPheLysGlyGlnLysArgIleLysCysPheAsnCysGlyLysGlu
GATGCAGAGAGGTAATTTTAAGGGCCAGAAAAGAATTAAGTGTTTCAACTGTGGCAAAGA

 GlyHisLeuAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLys
AGGACACCTAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAA
                                     1600   →POL
                                            PhePheArgGluAsnLeu
 GluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrp
GGAAGGACACCAAATGAAAGACTGCACTGAGAGACAGGCTAATTTTTTAGGGAAAATTTG

AlaPheProGlnGlyLysAlaArgGluPheProSerGluGlnThrArgAlaAsnSerPro
 ProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPro
GCCTTCCCACAAGGGAAGGCCAGGGAATTTCCTTCAGAGCAGACCAGAGCCAACAGCCCC
                 1700
ThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeuSerGluThrGlyAlaGlu
 ProAlaGluSerPheGlyPheGlyGluGluIleLysProSerGlnLysGlnGluGlnLys
ACCAGCAGAGAGCTTCGGGTTTGGGGAGGAGATAAAACCCTCTCAGAAACAGGAGCAGAA
                                                        1800
ArgGlnGlyIleValSerPheSerPheProGlnIleThrLeuTrpGlnArgProValVal
 AspLysGluLeuTyrProLeuAlaSerLeuLysSerLeuPheGlyAsnAspGlnLeuSer
AGACAAGGAATTGTATCCTTTAGCTTCCCTCAAATCACTCTTTGGCAACGACCAGTTGTC
GAG←
ThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThr
 Gln
ACAGTAAGAGTAGGAGGACAGCTAAAAGAAGCTCTATTAGACACAGGAGCAGATGATACA
                                     1900
ValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGly
GTATTAGAAGAAATAAATTTGCCAGGAAAATGGAAACCAAAAATGATAGGGGGAATTGGA

GlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAla
GGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAGGCT
                2000
```

FIG. 7B

```
IleGlyThrIleLeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThr
ATAGGTACAATATTGGTAGGACCTACACCTGTCAACATAATTGGACGAAATATGTTGACT
                                                        2100
GlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeu
CAGATTGGTTGTACTTTAAATTTTCCAATTAGTCCTATTGAGACTGTACCAGTAAAATTA

LysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGluGluLysIleLys
AAGCCAGGGATGGATGGCCCAAGGGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAA
                                    2200
AlaLeuThrGluIleCysLysAspMetGluLysGluGlyLysIleLeuLysIleGlyPro
GCATTAACAGAAATTTGTAAAGATATGGAAAAGGAAGGAAAAATTTTAAAAATTGGGCCT

GluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArg
GAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGCACTAAATGGAGA
                2300
LysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeu
AAATTAGTGAATTTCAGAGAGCTTAATAAAAGAACTCAAGATTTTTGGGAAGTTCAATTA
                                                        2400
GlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGly
GGAATACCACATCCTGCTGGGTTGAAAAAGAAAAAATCAGTCACAGTATTGGATGTGGGG

AspAlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIle
GATGCATATTTTTCAGTCCCTTTAGATGAAGATTTCAGGAAGTATACTGCATTCACTATA
                                    2500
ProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGly
CCCAGTATTAATAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTACCACAGGGA

TrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArg
TGGAAAGGATCACCAGCAATATTCCAGAGTAGCATGACAAAAATCTTAGAACCCTTTAGA
                2600
ThrLysAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAsp
ACAAAAAATCCAGAAATAGTCATATACCAATACATGGATGATTTGTATGTAGGGTCTGAT
                                                        2700
LeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGluHisLeuLeuLysTrp
TTAGAAATAGGACAACATAGAACAAAAATAGAGGAACTAAGAGAACATCTATTGAAATGG

GlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyr
GGATTTACCACACCAGACAAAAAGCATCAGAAAGAACCCCCATTTCTTTGGATGGGGTAT
                                    2800
GluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysGluSerTrp
GAACTCCACCCTGACAAATGGACAGTGCAGCCTATACAACTGCCAGACAAGGAAAGCTGG

ThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrPro
ACTGTCAATGATATACAGAAATTGGTGGGAAAACTAAATTGGGCAAGTCAGATTTATCCA
                2900
GlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIle
GGAATTAAAGTAAAGCAATTATGTAAACTCCTTAGGGGAGCAAAAGCACTAACAGACATA
                                                        3000
ValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGlu
GTACCATTAACTGCAGAGGCAGAATTAGAATTGGCAGAGAACAGGGAAATTCTAAAAGAA
```

FIG. 7C

```
ProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGln
CCAGTGCATGGGGTATATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAG
                                        3100
GlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyrLysAsnLeuLysThrGly
GGGCAAGGTCAATGGACATATCAAATATACCAAGAGCAATATAAAAATCTGAAAACAGGG

LysTyrAlaArgIleLysSerAlaHisThrAsnAspValLysGlnLeuThrGluAlaVal
AAGTATGCAAGAATAAAGTCTGCCCACACTAATGATGTAAAACAATTAACAGAAGCAGTG
                    3200
GlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArgLeuPro
CAAAAGATAGCCCAAGAAAGCATAGTAATATGGGGAAAAACTCCTAAATTTAGACTACCC
                                                        3300
IleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGlnAlaThrTrpIlePro
ATACAAAAAGAAACATGGGAGGCATGGTGGACAGAATATTGGCAAGCCACCTGGATCCCT

GluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGlu
GAATGGGAGTTTGTCAATACTCCTCCCCTAGTAAAACTATGGTACCAGTTAGAAACAGAA
                                        3400
ProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLys
CCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGAGAAACTAAAAAG

GlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSerLeuThrGluThr
GGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAGGTTGTCTCCTTAACTGAAACA
                    3500
ThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlySerGlu
ACAAATCAGAAGACTGAATTACAAGCAATCCACTTAGCTTTACAGGATTCAGGATCAGAA
                                                        3600
ValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLys
GTAAACATAGTAACAGACTCACAGTATGCATTAGGGATTATTCAAGCACAACCAGATAAA

SerGluSerGluIleValAsnGlnIleIleGluGlnLeuIleGlnLysAspLysValTyr
AGTGAATCAGAGATTGTTAATCAAATAATAGAGCAATTAATACAGAAGGACAAGGTCTAC
                                        3700
LeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuVal
CTGTCATGGGTACCAGCACACAAAGGGATTGGAGGAAATGAACAAGTAGATAAATTAGTC

SerSerGlyIleArgLysValLeuPheLeuAspGlyIleAspLysAlaGlnGluGluHis
AGCAGTGGAATCAGAAAGGTACTATTTTTAGATGGGATAGATAAGGCTCAAGAAGAACAT
                    3800
GluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeuProProIleVal
GAAAAATATCACAGCAATTGGAGAGCAATGGCTAGTGACTTTAATCTACCACCTATAGTA
                                                        3900
AlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluAlaMetHisGly
GCGAAGGAAATAGTAGCCAGCTGTGATAAATGTCAACTAAAAGGGGAAGCCATGCATGGA

GlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeuGluGlyLysIle
CAAGTAGACTGTAGTCCAGGGATATGGCAATTAGATTGCACACATCTAGAAGGAAAAATA
                                        4000
IleIleValAlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGlu
ATCATAGTAGCAGTCCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATCCCAGCAGAA

ThrGlyGlnGluThrAlaTyrPheIleLeuLysLeuAlaGlyArgTrpProValLysVal
ACAGGACAGGAGACAGCATACTTTATACTAAAATTAGCAGGAAGATGGCCAGTAAAAGTA
                    4100
```

FIG. 7D

```
ValHisThrAspAsnGlySerAsnPheThrSerAlaAlaValLysAlaAlaCysTrpTrp
GTACACACAGACAATGGCAGCAATTTCACCAGTGCTGCAGTTAAAGCAGCCTGTTGGTGG
                                                        4200
AlaAsnIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGlnGlyValValGlu
GCAAATATCAAACAGGAATTTGGAATTCCCTACAACCCCCAAAGTCAAGGAGTAGTGGAA

SerMetAsnLysGluLeuLysLysIleIleGlyGlnValArgGluGlnAlaGluHisLeu
TCTATGAATAAGGAATTAAAGAAAATCATAGGGCAGGTAAGAGAGCAAGCTGAACACCTT
                                        4300
LysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLysGlyGlyIleGly
AAGACAGCAGTACAAATGGCAGTGTTCATTCACAATTTTAAAAGAAAAGGGGGGATTGGG

GlyTyrSerAlaGlyGluArgIleIleAspMetIleAlaThrAspIleGlnThrLysGlu
GGGTACAGTGCAGGGGAAAGAATAATAGACATGATAGCAACAGACATACAAACTAAAGAA
                       4400
LeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArgAspAsnArgAsp
TTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAACAGAGAC
                                                        4500
ProIleTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaValValIleGln
CCAATTTGGAAAGGACCAGCAAAACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACAG
                                                         →Q
AspAsnSerAspIleLysValValProArgArgLysAlaLysIleIleArgAspTyrGly
                                                         MetGlu
GACAATAGTGATATAAAGGTAGTACCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGA
                                        4600     POL←
LysGlnMetAlaGlyAspAspCysValAlaGlyGlyGlnAspGluAsp
 AsnArgTrpGlnValMetIleValTrpGlnValAspArgMetArgIleArgThrTrpHis
AAACAGATGGCAGGTGATGATTGTGTGGCAGGTGGACAGGATGAGGATTAGAACATGGCA

 SerLeuValLysHisHisMetTyrValSerLysLysAlaLysAsnTrpPheTyrArgHis
CAGTTTAGTAAAACATCATATGTATGTCTCAAAGAAAGCTAAAAATTGGTTTTATAGACA
                       4700
 HisTyrGluSerArgHisProLysValSerSerGluValHisIleProLeuGlyAspAla
TCACTATGAAAGCAGGCATCCAAAAGTAAGTTCAGAAGTACACATCCCACTAGGGGATGC
                                                        4800
 ArgLeuValValArgThrTyrTrpGlyLeuGlnThrGlyGluLysAspTrpHisLeuGly
TAGATTAGTAGTAAGAACATATTGGGGTCTGCAAACAGGAGAAAAAGACTGGCACTTGGG

 HisGlyValSerIleGluTrpArgGlnLysArgTyrSerThrGlnLeuAspProAspLeu
TCATGGGGTCTCCATAGAATGGAGGCAGAAAAGATATAGCACACAACTAGATCCTGACCT
                                        4900
 AlaAspGlnLeuIleHisLeuTyrTyrPheAspCysPheSerGluSerAlaIleArgGln
AGCAGACCAACTGATTCATCTGTACTATTTTGATTGTTTTTCAGAATCTGCCATAAGACA

 AlaIleLeuGlyHisIleValSerProArgCysAspTyrGlnAlaGlyHisAsnLysVal
AGCCATATTAGGACATATAGTTAGTCCTAGGTGTGATTATCAAGCAGGACATAACAAGGT
                       5000
 GlySerLeuGlnTyrLeuAlaLeuThrAlaLeuIleAlaProLysLysThrArgProPro
AGGATCTTTACAGTATTTGGCACTAACAGCATTAATAGCACCAAAAAAGACAAGGCCACC
                                   →R                   5100
                                   MetGluGlnAlaProAlaAspGlnGly
 LeuProSerValArgLysLeuThrGluAspArgTrpAsnLysProGlnGlnThrLysGly
TTTGCCTAGTGTTAGGAAGCTAACAGAAGATAGATGGAACAAGCCCCAGCAGACCAAGGG
```

FIG. 7E

Q←

ProGlnArgGluProHisAsnGluTrpThrLeuGluLeuLeuGluGluLeuLysGlnGlu
HisArgGlySerHisThrMetAsnGlyHis
CCACAGAGGGAGCCACACAATGAATGGACATTAGAACTTTTAGAGGAGCTTAAGCAAGAA
5200

AlaValArgHisPheProArgIleTrpLeuHisSerLeuGlyGlnHisIleTyrGluThr
GCTGTCAGACACTTTCCTAGGATATGGCTCCATAGTTTAGGACAACATATCTATGAAACT

TyrGlyAspThrTrpGluGlyValGluAlaIleIleArgSerLeuGlnGlnLeuLeuPhe
TATGGGGATACCTGGGAAGGAGTTGAAGCTATAATAAGAAGTCTGCAACAACTGCTGTTT
5300

IleHisPheArgIleGlyCysGlnHisSerArgIleGlyIleThrArgGlnArgArgAla
ATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTACTCGACAGAGAAGAGCA
5400

→S R←

ArgAsnGlySerSerArgSer
MetAspProValAspProAsnLeuGluProTrpAsnHisProGlySerGlnProArg
AGAAATGGATCCAGTAGATCCTAACTTAGAGCCCTGGAACCATCCAGGGAGTCAGCCTAG

ThrProCysAsnLysCysTyrCysLysLysCysCysTyrHisCysGlnMetCysPheIle
GACGCCTTGTAATAAGTGTTATTGTAAAAAGTGCTGCTATCATTGCCAAATGTGCTTCAT
5500

ThrLysGlyLeuGlyIleSerTyrGlyArgLysLysArgArgGlnArgArgArgProPro
AACGAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCC

S←

GlnGlyAsnGlnAlaHisGlnAspProLeuProGluGln
TCAGGGCAATCAGGCTCATCAAGATCCTCTACCAGAGCAGTAAGTAGTATATGTAATACA
5600

ACCTTTAGTGATATTAGCAATAGTAGCATTAGTAGTAACGCTAATAATAGCAATAGTTGT
5700

GTGGACCATAGTATTTATAGAAATTAGGAAAATAAGAAGACAAAGGAAAATAGACAGGTT

→ENV

MetArgValArgGluIleGlnArg
GATTGATAGAATAAGAGAAAGAGCAGAAGATAGTGGCAATGAGAGTGAGGGAGATACAGA
5800

AsnTyrGlnAsnTrpTrpArgTrpGlyMetMetLeuLeuGlyMetLeuMetThrCysSer
GGAATTATCAAAACTGGTGGAGATGGGGCATGATGCTCCTTGGGATGTTGATGACCTGTA

IleAlaGluAspLeuTrpValThrValTyrTyrGlyValProValTrpLysGluAlaThr
GTATTGCAGAAGATTTGTGGGTTACAGTTTATTATGGGGTACCTGTGTGGAAAGAAGCAA
5900

ThrThrLeuPheCysAlaSerAspAlaLysSerTyrGluThrGluValHisAsnIleTrp
CCACTACTCTATTTTGTGCATCAGATGCTAAATCATATGAAACAGAAGTACATAACATCT
6000

AlaThrHisAlaCysValProThrAspProAsnProGlnGluIleGluLeuGluAsnVal
GGGCTACACATGCCTGTGTACCCACGGACCCCAACCCACAAGAAATAGAACTGGAAAATG

ThrGluGlyPheAsnMetTrpLysAsnAsnMetValGluGlnMetHisGluAspIleIle
TCACAGAAGGGTTTAACATGTGGAAAAATAACATGGTGGAGCAGATGCATGAGGATATAA
6100

FIG. 7F

SerLeuTrpAspGlnSerLeuLysProCysValLysLeuThrProLeuCysValThrLeu
TCAGTTTATGGGATCAAAGCCTAAAACCATGTGTAAAGCTAACCCCACTCTGTGTCACTT

AsnCysThrAsnValAsnGlyThrAlaValAsnGlyThrAsnAlaGlySerAsnArgThr
TAAACTGCACTAATGTGAATGGGACTGCTGTGAATGGGACTAATGCTGGGAGTAATAGGA
6200
AsnAlaGluLeuLysMetGluIleGlyGluValLysAsnCysSerPheAsnIleThrPro
CTAATGCAGAATTGAAAATGGAAATTGGAGAAGTGAAAAACTGCTCTTTCAATATAACCC
6300
ValGlySerAspLysArgGlnGluTyrAlaThrPheTyrAsnLeuAspLeuValGlnIle
CAGTAGGAAGTGATAAAAGGCAAGAATATGCAACTTTTTATAACCTTGATCTAGTACAAA

AspAspSerAspAsnSerSerTyrArgLeuIleAsnCysAsnThrSerValIleThrGln
TAGATGATAGTGATAATAGTAGTTATAGGCTAATAAATTGTAATACCTCAGTAATTACAC
6400
AlaCysProLysValThrPheAspProIleProIleHisTyrCysAlaProAlaGlyPhe
AGGCTTGTCCAAAGGTAACCTTTGATCCAATTCCCATACATTATTGTGCCCCAGCTGGTT

AlaIleLeuLysCysAsnAspLysLysPheAsnGlyThrGluIleCysLysAsnValSer
TTGCAATTCTAAAGTGTAATGATAAGAAGTTCAATGGAACGGAAATATGTAAAAATGTCA
6500
ThrValGlnCysThrHisGlyIleLysProValValSerThrGlnLeuLeuLeuAsnGly
GTACAGTACAATGTACACATGGAATTAAGCCAGTGGTGTCAACTCAACTGCTGTTAAATG
6600
SerLeuAlaGluGluGluIleMetIleArgSerGluAsnLeuThrAspAsnThrLysAsn
GCAGTCTAGCAGAAGAAGAGATAATGATTAGATCTGAAAATCTCACAGACAATACTAAAA

IleIleValGlnLeuAsnGluThrValThrIleAsnCysThrArgProGlyAsnAsnThr
ACATAATAGTACAGCTTAATGAAACTGTAACAATTAATTGTACAAGGCCTGGAAACAATA
6700
ArgArgGlyIleHisPheGlyProGlyGlnAlaLeuTyrThrThrGlyIleValGlyAsp
CAAGAAGAGGGATACATTTCGGCCCAGGGCAAGCACTCTATACAACAGGGATAGTAGGAG

IleArgArgAlaTyrCysThrIleAsnGluThrGluTrpAspLysThrLeuGlnGlnVal
ATATAAGAAGAGCATATTGTACTATTAATGAAACAGAATGGGATAAAACTTTACAACAGG
6800
AlaValLysLeuGlySerLeuLeuAsnLysThrLysIleIlePheAsnSerSerSerGly
TAGCTGTAAAACTAGGAAGCCTTCTTAACAAAACAAAAATAATTTTTAATTCATCCTCAG
6900
GlyAspProGluIleThrThrHisSerPheAsnCysArgGlyGluPhePheTyrCysAsn
GAGGGGACCCAGAAATTACAACACACAGTTTTAATTGTAGAGGGGAATTTTTCTACTGTA

ThrSerLysLeuPheAsnSerThrTrpGlnAsnAsnGlyAlaArgLeuSerAsnSerThr
ATACATCAAAACTGTTTAATAGTACATGGCAGAATAATGGTGCAAGACTAAGTAATAGCA
7000
GluSerThrGlySerIleThrLeuProCysArgIleLysGlnIleIleAsnMetTrpGln
CAGAGTCAACTGGTAGTATCACACTCCCATGCAGAATAAAACAAATTATAAATATGTGGC

LysThrGlyLysAlaMetTyrAlaProProIleAlaGlyValIleAsnCysLeuSerAsn
AGAAAACAGGAAAAGCTATGTATGCCCCTCCCATCGCAGGAGTCATCAACTGTTTATCAA
7100
IleThrGlyLeuIleLeuThrArgAspGlyGlyAsnSerSerAspAsnSerAspAsnGlu
ATATTACAGGGCTGATATTAACAAGAGATGGTGGAAATAGTAGTGACAATAGTGACAATG
7200

FIG. 7G

```
 ThrLeuArgProGlyGlyGlyAspMetArgAspAsnTrpIleSerGluLeuTyrLysTyr
AGACCTTAAGACCTGGAGGAGGAGATATGAGGGACAATTGGATAAGTGAATTATATAAAT

 LysValValArgIleGluProLeuGlyValAlaProThrLysAlaLysArgArgValVal
ATAAAGTAGTAAGAATTGAACCCCTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGG
                                        7300
 GluArgGluLysArgAlaIleGlyLeuGlyAlaMetPheLeuGlyPheLeuGlyAlaAla
TGGAAAGAGAAAAAAGAGCAATAGGACTAGGAGCCATGTTCCTTGGGTTCTTGGGAGCAG

 GlySerThrMetGlyAlaAlaSerLeuThrLeuThrValGlnAlaArgGlnLeuLeuSer
CAGGAAGCACGATGGGCGCAGCGTCACTAACGCTGACGGTACAGGCCAGACAGTTACTGT
                    7400
 GlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGlnHisLeuLeu
CTGGTATAGTGCAACAGCAAAACAATTTGCTGAGGGCTATAGAGGCGCAACAGCATCTGT
                                                        7500
 GlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaValGluArgTyr
TGCAACTCACGGTCTGGGGCATTAAACAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGAT

 LeuGlnAspGlnArgLeuLeuGlyMetTrpGlyCysSerGlyLysHisIleCysThrThr
ACCTACAGGATCAACGGCTCCTAGGAATGTGGGGTTGCTCTGGAAAACACATTTGCACCA
                                        7600
 PheValProTrpAsnSerSerTrpSerAsnArgSerLeuAspAspIleTrpAsnAsnMet
CATTTGTGCCTTGGAACTCTAGTTGGAGTAATAGATCTCTAGATGACATTTGGAATAATA

 ThrTrpMetGlnTrpGluLysGluIleSerAsnTyrThrGlyIleIleTyrAsnLeuIle
TGACCTGGATGCAGTGGGAAAAAGAAATTAGCAATTACACAGGCATAATATACAACTTAA
                    7700
 GluGluSerGlnIleGlnGlnGluLysAsnGluLysGluLeuLeuGluLeuAspLysTrp
TTGAAGAATCGCAAATCCAGCAAGAAAAGAATGAAAAGGAATTATTGGAATTGGACAAGT
                                                        7800
 AlaSerLeuTrpAsnTrpPheSerIleSerLysTrpLeuTrpTyrIleArgIlePheIle
GGGCAAGTTTGTGGAATTGGTTTAGCATATCAAAATGGCTGTGGTATATAAGAATATTCA

 IleValValGlyGlyLeuIleGlyLeuArgIleIlePheAlaValLeuSerLeuValAsn
TAATAGTAGTAGGAGGCTTAATAGGTTTAAGAATAATTTTTGCTGTGCTTTCTTTAGTAA
                                        7900
 ArgValArgGlnGlyTyrSerProLeuSerLeuGlnThrLeuLeuProThrProArgGly
ATAGAGTTAGGCAGGGATACTCACCTCTGTCGTTGCAGACCCTCCTCCCAACACCGAGGG

 ProProAspArgProGluGlyIleGluGluGluGlyGlyGluGlnGlyArgGlyArgSer
GACCACCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGCAAGGCAGAGGCAGAT
                    8000
 IleArgLeuValAsnGlyPheSerAlaLeuIleTrpAspAspLeuArgAsnLeuCysLeu
CAATTCGATTGGTGAACGGATTCTCAGCACTTATCTGGGACGACCTGAGGAACCTGTGCC
                                                        8100
 PheSerTyrHisArgLeuArgAspLeuLeuLeuIleAlaThrArgIleValGluLeuLeu
TCTTCAGTTACCACCGCTTGAGAGACTTACTCTTAATTGCAACGAGGATTGTGGAACTTC

 GlyArgArgGlyTrpGluAlaLeuLysTyrLeuTrpAsnLeuLeuGlnTyrTrpGlyGln
TGGGACGCAGGGGGTGGGAAGCCCTCAAATATCTGTGGAATCTCCTGCAATATTGGGGTC
                                        8200
```

FIG. 7H

GluLeuLysAsnSerAlaIleSerLeuLeuAsnThrThrAlaIleAlaValAlaGluCys
AGGAACTGAAGAATAGTGCTATTAGCTTGCTTAATACCACAGCAATAGCAGTAGCTGAAT

ThrAspArgValIleGluIleGlyGlnArgPheGlyArgAlaIleLeuHisIleProArg
GCACAGATAGGGTTATAGAAATAGGACAAAGATTTGGTAGAGCTATTCTCCACATACCTA
8300

EW← →F
MetGlyGlyLysTrpSerLys
ArgIleArgGlnGlyPheGluArgAlaLeuLeu
GAAGAATTAGACAGGGCTTCGAAAGGGCTTTGCTATAACATGGGTGGCAAGTGGTCAAAA
8400

SerSerIleValGlyTrpProLysIleArgGluArgIleArgArgThrProProThrGlu
AGTAGCATAGTAGGATGGCCTAAGATTAGGGAAAGAATAAGACGAACTCCCCCAACAGAA

ThrGlyValGlyAlaValSerGlnAspAlaValSerGlnAspLeuAspLysCysGlyAla
ACAGGAGTAGGAGCAGTATCTCAAGATGCAGTATCTCAAGATTTAGATAAATGTGGAGCA
8500

AlaAlaSerSerSerProAlaAlaAsnAsnAlaSerCysGluProProGluGluGluGlu
GCCGCAAGCAGCAGTCCAGCAGCTAATAATGCTAGTTGTGAACCACCAGAAGAAGAGGAG

GluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLysGlyAla
GAGGTAGGCTTTCCAGTCCGTCCTCAGGTACCTTTAAGACCAATGACTTATAAAGGAGCT
8600 →U3

PheAspLeuSerHisPheLeuLysGluLysGlyGlyLeuAspGlyLeuValTrpSerPro
TTTGATCTCAGCCACTTTTTAAAAGAAAAGGGGGGACTGGATGGGTTAGTTTGGTCCCCA
8700

LysArgGlnGluIleLeuAspLeuTrpValTyrHisThrGlnGlyTyrPheProAspTrp
AAAAGACAAGAAATCCTTGATCTGTGGGTCTACCACACACAAGGCTACTTCCCTGATTGG

GlnAsnTyrThrProGlyProGlyIleArgPheProLeuThrPheGlyTrpCysPheLys
CAGAATTACACACCAGGGCCAGGGATTAGATTCCCACTGACCTTCGGATGGTGCTTTAAG
8800

LeuValProMetSerProGluGluValGluGluAlaAsnGluGlyGluAsnAsnCysLeu
TTAGTACCAATGAGTCCAGAGGAAGTAGAGGAGGCCAATGAAGGAGAGAACAACTGTCTG

LeuHisProIleSerGlnHisGlyMetGluAspAlaGluArgGluValLeuLysTrpLys
TTACACCCTATTAGCCAACATGGAATGGAGGACGCAGAAAGAGAAGTGCTAAAATGGAAG
8900

PheAspSerSerLeuAlaLeuArgHisArgAlaArgGluGlnHisProGluTyrTyrLys
TTTGACAGCAGCCTAGCACTAAGACACAGAGCCAGAGAACAACATCCGGAGTACTACAAA
F← 9000

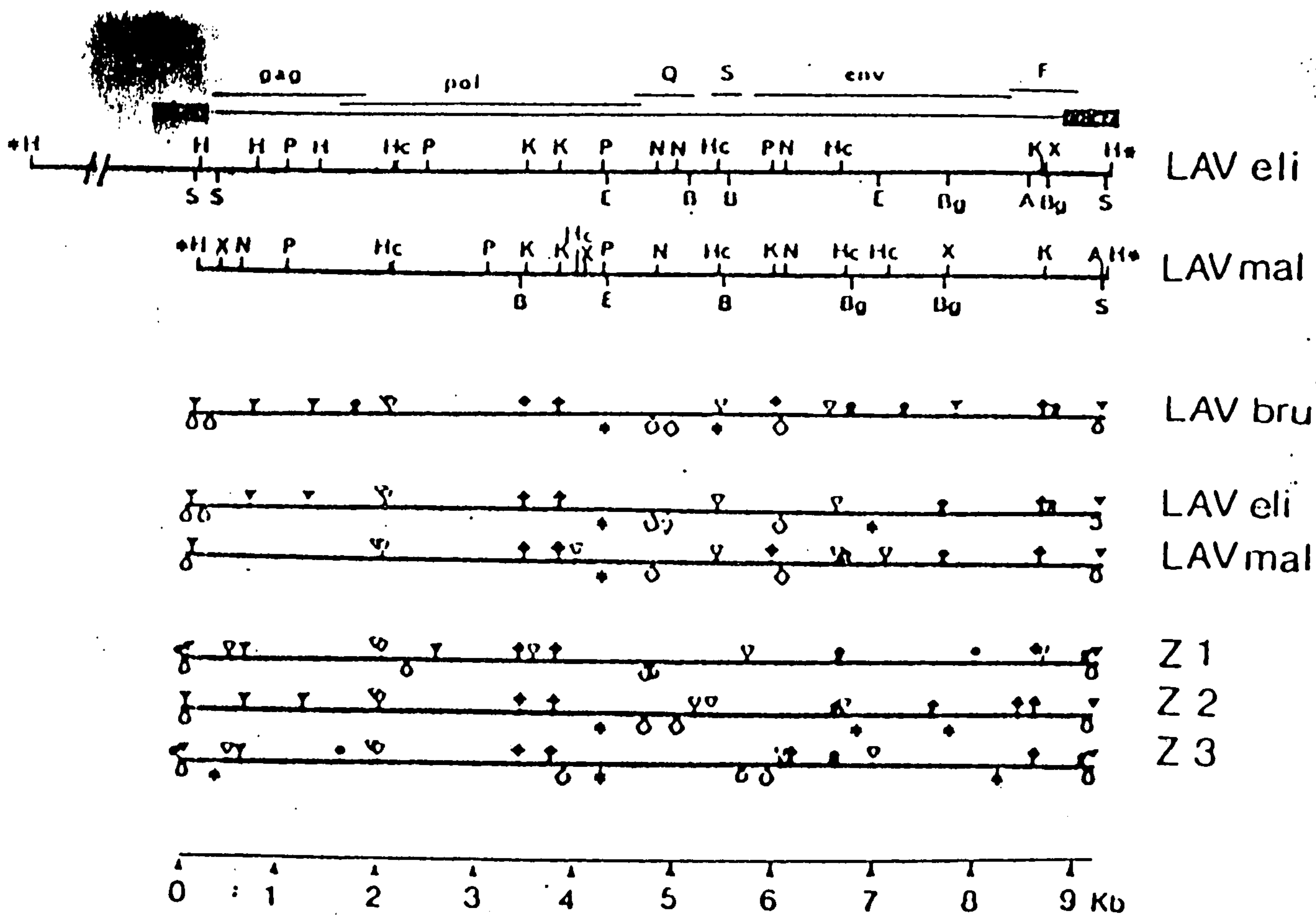
AspCys
GACTGCTGACACAGAAGTTGCTGACAGGGGACTTTCCGCTGGGGACTTTCCAGGGGAGGC

GTAACTTGGGCGGGACCGGGGAGTGGCTAACCCTCAGATGCTGCATATAAGCAGCTGCTT
U3← →R 9100

TTCGCCTGTACTGGGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGC

AAGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAA
9200

FIG. 7I

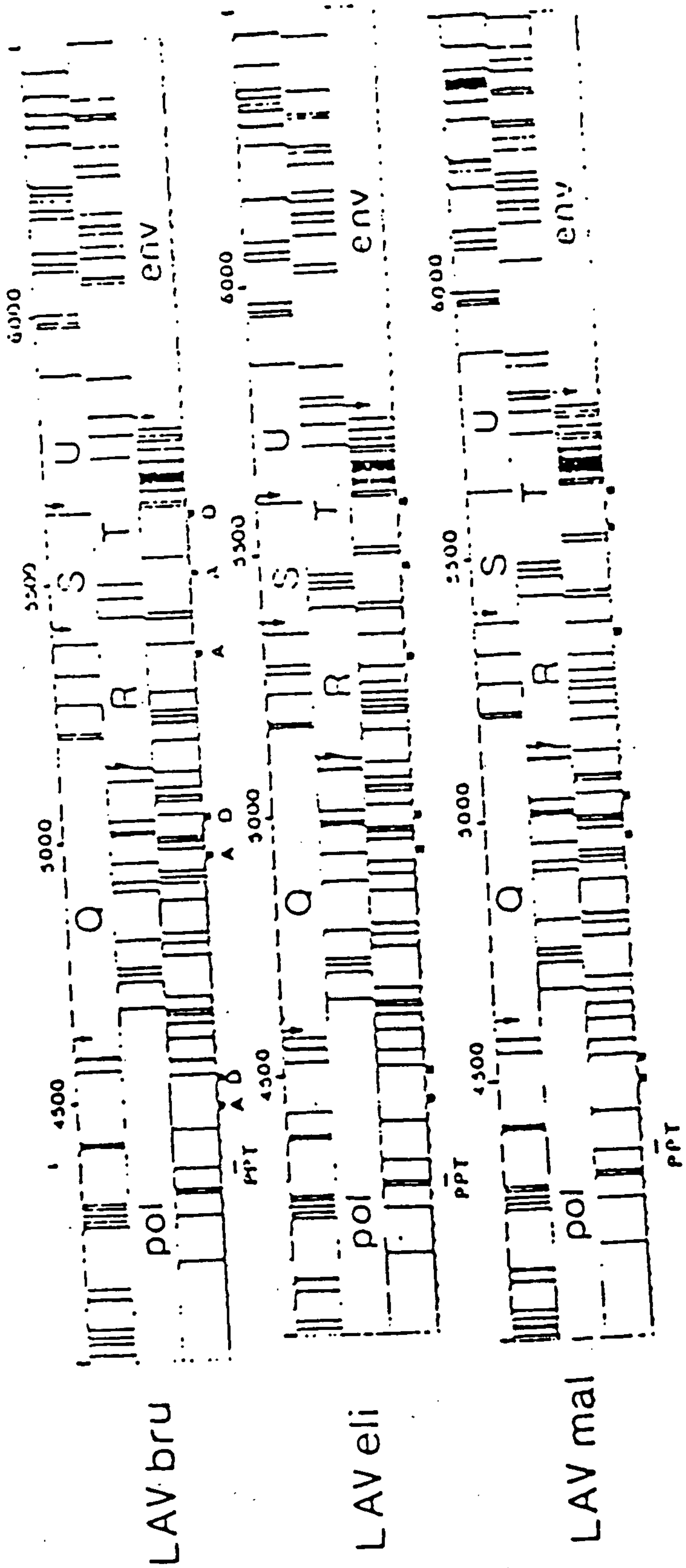


gag
pol
Q
S
env
F
LAV eli
LAV mal
LAV bru
LAV eli
LAV mal
Z 1
Z 2
Z 3
0
1
2
3
4
5
6
7
8
9 Kb

FIG. 1A

FIG. 1B

FIG. 2


LAV bru
LAV eli
LAV mal
pol
env
Q
R
S
T
U
4500
5000
5500
6000
PPT
A
D

86ETE0"SZ6Th060

Coll
66

GAG

```
                     10         20         30         40         50         60         70         80
LAV BRU     HCARASVLSC CELDRHEKIR LRPCCKKKYK LKHIVWASRE LERFAVHPCL LETSECCRQI LCQLQPSLQT CSELLRSLYH
ARV 2                      K
LAV HAL                 K  A               R    L            L        C   Q   HE   ST K      IK
LAV ELI                 K  K               R               Y L            K   I     AI    T
                                                                           |p25
                     90        100        110        120        130        140        150        160
LAV BRU     TVATLYCVHQ RIEIKDTKEA LDKIEEEQHK SKKKAQQAAA ------DTCH SSQVSQHYPI VQHIQCQHVH QAISPRTLHA
ARV 2                    DV        E                    ----AAC  H               L
LAV HAL                  DV        .    I     RQ T      AQQAAAA KH   S           A    I
LAV ELI              K G DV        E H                  ------   H N             L

                    170        180        190        200        210        220        230        240
LAV BRU     VVKVVEEKAF SPEVIPHFSA LSECATPQDL HTHLHTVGGH QAAHQHLKET IHEEAAEWDK VHPVHACPIA PCQHREPRGS
ARV 2
LAV HAL          I                            H   I             D        D             P
LAV ELI          I                                                            L

                    250        260        270        280        290        300        310        320
LAV BRU     DIAGTTSTLQ EQIGWHTHHP PIPVGEIYKR WIILGLNKIV RHYSPTSILD IRQGPKEPFR DYVDRFYKTL RAEQASQEVK
ARV 2                                                                                                D
LAV HAL                        S       D                      V                      F         T
LAV ELI                    A   S                 V            V                                  D
                                                                                  |p13
                    330        340        350        360        370        380        390        400
LAV BRU     KWHTETLLVQ HAHPDCKTIL KALGPAATLE EHHTACQGVG GPGHKARVLA EAHSQVTHS- ATIHHQKGHF RHQRKIVKCF
ARV 2                                                                          P-  H               T
LAV HAL                                G                  S             A   T  A         KC - KI
LAV ELI                                Q                  S             A   V T A        KCP   I

                    410        420        430        440        450        460        470        480
LAV BRU     NCGKEGHIAR HCRAPRKKGC WKCGKEGHQH KDCTERQAHF LCKIWPSYKG RPGHFLQSRP EPTAPPFLQS RPEPTAPPEE
ARV 2                K             R  R                                             ---- --------
LAV HAL            L                                           H                    ---- ---------A
LAV ELI              K             R       L              R    H                    ---- ---------A

                    490        500        510
LAV BRU     SFRSGVETTT PSQKQEPIDK ELYPLTSLKS LFGNDPSSQ
ARV 2          F E K
LAV HAL       GF E IK-       QK        A  K       QL
LAV ELI       GF E I -       QK           K        L
```

FIG. 3A

Central region :

Q

```
                    10         20         30         40         50         60         70         80
LAV BRU    HENRWQVMIV WQVDRHRIRT WKSLVKHHMY VSCKARGWFY RHHYESPHPR ISSEVHIPLG DARLVITTYW GLHTGERDWH
ARV 2                                       I K  K          T     V            K              E
LAV MAL                              H        K  KH         R  K  V              VR       Q  K
LAV ELI       I                K              K  HK               K         E      K          E

                    90        100        110        120        130        140        150        160
LAV BRU    LGQGVSIEWR KKRYSTQVDP ELADQLIHLY YFDCFSDSAI RKALLGHIVS PRCEYQAGHN KVGSLQYLAL AALITPKKIK
ARV 2           A        K        G        H         E     KH I   YR                          T
LAV MAL       N         Q        L  D                  E    Q I        D              T   A    TR
LAV ELI                  R          G        N          E       I  D                  T   A   Q

                   170        180        190
LAV BRU    PPLPSVTKLT EDRWNKPQKT KGHRGSHTMN GH
ARV 2              K
LAV MAL            R          Q
LAV ELI            R          Q  R
```

R

```
                    10         20         30         40         50         60         70         80
LAV BRU    HEQAPEDQGP QREPHNEWTL ELLEELKNEA VRHFPRIWLH GLGQHIYETY GDTWAGVEAI IRILQQLLFI HFRIGCRHSR
ARV 2                      Y             R          P  S  Y                                   Q
LAV MAL         A                        Q             S            E          S              Q
LAV ELI         A          Y    A        S             S            V                         Q

                    90
LAV BRU    IGVTQQRRAR -NGASRS
ARV 2       II         R
LAV MAL     I R        -  S
LAV ELI     IIR        -  S
```

S (tat)

```
                    10         20         30         40         50         60         70
LAV BRU    HEPVDPRLEP WKHPGSQPKT ACTTCYCKKC CFHCQVCFTT KALGISYGRK KRRQRRRPPQ GSQTHQVSLS KQ
ARV 2              H          R    HH          YA  R  G               A     D     A
LAV MAL     D      N    N     R  P HK        Y  H  I  G                    H A  DP P E
LAV ELI     D      N    N     R  P HK H      Y P   LH G              G     G A   PIP
```

FIG. 3B

FIG. 3C

POL

```
                  10         20         30         40         50         60         70         80
LAV BRU   FFREDLAFLQ GKAREFSSEQ TRANSPTFSS EQTRANSPTR RELQVWGRDN NSLSEAGADR QCTVSFNFPQ ITLWQRPLVT
ARV 2                           ---  ---------            GE
LAV HAL     H  P       P        ---  ---------S    R    C - KT   T  E     I  S          V
LAV ELI     H  P     G L PK     ---  ---------S    R      -  P  KT  E                     A

                  90        100        110        120        130        140        150        160
LAV BRU   IKIGGQLKEA LLDTGADDTV LEENSLPGRW KPKMIGGIGG FIKVRQYDQI LIEICGHKAI GTVLVGPTPV NIIGRNLLTQ
ARV 2      R                        N   K                       PV
LAV HAL   VRV                      IN   K                                K      I             M
LAV ELI                             M   K                       P        Q

                 170        180        190        200        210        220        230        240
LAV BRU   IGCTLNFPIS PIETVPVKLK PGMDGPKVKQ WPLTEEKIKA LVEICTEMEK EGKISKIGPE NPYNTPVFAI KKKDSTKWRK
ARV 2
LAV HAL                            R                T   KD       L
LAV ELI                                             T    D        R          I

                 250        260        270        280        290        300        310        320
LAV BRU   LVDFRELNKR TQDFWEVQLG IPHPAGLKKK KSVTVLDVGD AYFSVPLDED FRKYTAFTIP SINNETPGIR YQYNVLPQGW
ARV 2                                                         K
LAV HAL     K
LAV ELI                                                                 S

                 330        340        350        360        370        380        390        400
LAV BRU   KGSPAIFQSS MTKILEPFRK QNPDIVIYQY MDDLYVGSDL EIGQHRTKIE ELRQHLLRWG LTTPDKKHQK EPPFLWMGYE
ARV 2                                                                   F
LAV HAL                     T K  E                                  E   K   F
LAV ELI                          EH                              K  E       F R

                 410        420        430        440        450        460        470        480
LAV BRU   LHPDKWTVQP IVLPEKDSWT VNDIQKLVGK LNWASQIYPG IKVRQLCKLL RGTKALTEVI PLTEEAELEL AENREILKEP
ARV 2                N                              A     K
LAV HAL              Q  D E                               K       A     DIV    A
LAV ELI            S K    E          N  ER

                 490        500        510        520        530        540        550        560
LAV BRU   VHGVYYDPSK DLIAEIQKQG QGQWTYQIYQ EPFKNLKTGK YARTRGAHTN DVKQLTEAVQ KITTESIVIW GKTPKFKLPI
ARV 2       E          V                                N                 VS          I
LAV HAL                                   QY            IKS                 AQ          R
LAV ELI                          N                      N              A    R S      R    R
```

09041925.031398

FIG. 3D

```
                570        580        590        600        610        620        630        640
LAV BRU  QKETWETWWT EYWQATWIPE WEFVNTPPLV KLWYQLEKEP IVGAETFYVD GAASRETKLG KAGYVTNRGR QKVVTLTDTT
ARV 2      I   A  H                                                N             D        SIA
LAV HAL        A                                 T                 N    K        D        S  E
LAV ELI           A                                   I            N             D        P

                650        660        670        680        690        700        710        720
LAV BRU  HQKTELQAIH LALQDSGLEV NIVTDSQYAL GIIQAQPDKS ESELVNQIIE QLIKKEKVYL AWVPAHKGIG GNEQVDKLVS
ARV 2                                                     S
LAV HAL                    S                            I          Q D     S
LAV ELI           K

                730        740        750        760        770        780        790        800
LAV BRU  AGIRKVLFLD GIDKAQDEHE KYHSNWRAMA SDFNLPPVVA KEIVASCDKC QLKGEAMHGQ VDCSPGIWQL DCTHLEGKVI
ARV 2             N       E                                                                   I
LAV HAL  S                E                      I                                            I
LAV ELI  Q                E       N

                810        820        830        840        850        860        870        880
LAV BRU  LVAVHVASGY IEAEVIPAET GQETAYFLLK LAGRWPVKTI HTDNGSNFTS TTVKAACWWA GIKQEFGIPY NPQSQGVVES
ARV 2
LAV HAL  I                            I           VV            AA         N
LAV ELI                                           VV            AA

                890        900        910        920        930        940        950        960
LAV BRU  HMKELKKIIG QVRDQAEHLK TAVQMAVFIH NFKRKGGIGG YSAGERIVDI IATDIQTKEL QKQITKIQNF RVYYRDSRDP
ARV 2      N                                                                                KK
LAV HAL                 E                                    I H                             H
LAV ELI                                       RR            I                  I

                970        980        990       1000       1010
LAV BRU  LWKGPAKLLW KGEGAVVIQD NSDIKVVPRR KAKIIRDYGK QMAGDDCVAS RQDED
ARV 2
LAV HAL  I                                                    G G
LAV ELI  I                     K           V
```

```
ENV                                     SP                                       OMP
                     10         20         30          40         50         60         70         80
LAV BRU     MRVK---EKY QNLWRWGWKW GTMLLGILMI CSATEKLWVT VYYGVPVWKE ATTTLFCASD AKAYDTEVHN VWATHACVPT
ARV 2        K  GTRRN         --- - L  M                                        R
LAV MAL        RELQRH  NW     --- -H   H  T   IA D                                  S E        I
LAV ELI       ARGIEHH[C] HW K  --- -I      T    ADH                                   S E  A     I

                     90        100        110        120        130        140        150        160
LAV BRU     DPNPQEVVLV NVTENFNMWK NDMVEQMHED IISLWDQSLK PCVKLTPLCV SLKCTDL-GN ATNTNSSNTN SSSGEMMME-
ARV 2                C .           N     Q                         T N     - K        ---  ----KGKE I
LAV MAL            IE E      G     N                               T N  NVN T  V CTNACS  RTNA LK  I
LAV ELI            IA E            N                               T N  S  E--L RN CTNG NV  TTEEKG----

                    170        180        190        200        210        220        230        240
LAV BRU     KGEIKNCSFN ISTSIRGKVQ KEYAFFYKLD IIPIDNDTTS -----YTLTS CNTSVITQAC PKVSFEPIPI HYCAPAGFAI
ARV 2                     T    D I    N L RN VV    AS T TNYTN R IN   R                          T
LAV MAL     -  V          TPVGSD R  -   T  N  LVQ  DSDN  ----S K IN                     T D
LAV ELI     ---N          VT VLKD K QV   L  K    V      SST -NSTN K IN      A

                    250        260        270        280        290        300        310        320
LAV BRU     LKCNNKTFNG TGPCTNVSTV QCTHGIRPVV STQLLLNGSL AEEEVVIRSA NFTDNAKTII VQLNQSVEIN CTRPNNNTRK
ARV 2                  K               I                        D    N              E A
LAV MAL         D K      EI K           K                    IH     E   L    T N      ET T        G    R
LAV ELI        RD K                                           I     E   L N    N   AN    E  K T   A  YQ   Q

                    330        340        350        360        370        380        390        400
LAV BRU     SIRIQRGPGR AFVTIGK-IG NMRQAHCNIS RAKWNATLKQ IASKLREQFG NNKT-IIFKQ SSGGDPEIVT HSFNCGGEFF
ARV 2          Y --         N T RI   DI K          Q  N   E    VK              - V N            N        R
LAV MAL     C NF--     Q  LY T I-V  DI R Y T N ETE DK   Q   V V   GSLL- -   K   NS             T          R
LAV ELI     RTP --  L Q SLY TKS-RS IIG          Q SK   Q   V R   GTLL- -   I K  P              T

                    410        420        430        440        450        460        470        480
LAV BRU     YCNSTQLFNS TWFNSTWSTE GSNNTEGSDT ITLPCRIKQF INMWQEVGKA MYAPPISGQI RCSSNITGLL LTRDGGNN--
ARV 2          T     N    -----RLN NTEG K N    I         I                        G    S                  T -V
LAV MAL          TSK        Q NCAKL-  -   S  STGS              I       KT            A V    N L        I          NSSD
LAV ELI          TSG        NI A NNI  TES NSTNTN       Q       I    K VAGR-     I      ERN   L                   I --

                    490        500        510        520        530        540  TMP    550        560
LAV BRU     NNGSEIFRPG GGDMRDNWRS ELYKYKVVKI EPLGVAPTKA KRRVVQREKR AVGI-GALFL GFLGAAGSTM GARSMTLTVQ
ARV 2       T DT V                         I          I                              V  N                   V L
LAV MAL      SDN TL              I         R                          E         I L-   N                  A L
LAV ELI      STN T                         Q            R             E         I L-   N                     V
```

FIG. 3E

```
                570       580       590       600       610       620       630       640
LAV BRU   ARQLLSGIVQ QQNNLLRAIE AQQHLLQLTV WGIKQLQARI LAVERYLKDQ QLLGIWGCSG KLICTTAVPW NASWSNKSLE
ARV 2                                           V          R                H      P      S     R  D
LAV HAL                                         V          Q    R    H      H      N      S     R  N
LAV ELI       H

                650       660       670       680       690       700       710       720
LAV BRU   QIWNNMTWME WDREINNYTS LIHSLIEESQ NQQEKNEQEL LELDKWASLW NWFNITNWLW YIKIFIMIVG GLVGLRIVFA
                                                                 S
ARV 2     D  D   Q  E   D   N T YT L                             S  SK      R    IV      I     I
LAV HAL   C       Q  EK  S   G I YN        I       K             S  Q                 I      I
LAV ELI   E  Q       E   D   G   Y         T       K

                730       740       750       760       770       780       790       800
LAV BRU   VLSIVNRVRQ GYSPLSFQTH LPTPRGP-DR PEGIEEEGGE RDRDRSIRLV NGSLALIWDD LRSLCLFSYH RLRDLLLIVT
                                                                                              AA
ARV 2                          R   V   -    D                  V    D  F       E          R   A
LAV HAL      L              L  L       P              QG  G         FS           H        I  AV
LAV ELI      L                 L   A   -       T        G      V  L FS

                810       820       830       840       850       860       870
LAV BRU   RIVELLGRRG WEALKYWWNL LQYWSQELKN SAVSLLNATA IAVAECTDRV IEVVQGACRA IRHIPRRIRQ GLERILL
                                                                   A  R Y    L  H          L
ARV 2      T  I  N        S       I            W         T         IG  RPG    L         F  A
LAV HAL              L           G       I       T        [D]      II  N      VLN           S
LAV ELI         DI   L                R        S    FD I

F
                10        20        30        40        50        60        70        80
LAV BRU   MGGKWSKSSV VGWPTVRERM R----RAEPA ADGVGAASR- ----DLEKHG AITSSNTAAT NAACAWLEAQ EE-EEVGFPV
                                                      V  -  ----                    D         -
ARV 2          R  N  G   SAI      RAEP                                              S --- PP  E
LAV HAL          I       KI   I   ---- TP  T  ET      V QD  AVSQ   D  C   AA   SP  N           SD
LAV ELI          I       AI   I   ---- TN             V  -  ----                 S   D

                90        100       110       120       130       140       150       160
LAV BRU   TPQVPLRPMT YKAAVDLSHF LKEKGGLEGL IHSQRRQDIL DLWIYHTQGY FPDWQNYTPG PGVRYPLTFG WCYKLVPVEP
                                                 W     E                              I          F
ARV 2     R                 L I                 D VG  PK E       V                    I  F      F     NS
LAV HAL   R              G F                      W  KK E       V  N    I             I         E     D
LAV ELI   N              E L

                170       180       190       200       210
LAV BRU   DKVEEANKGE NTSLLHPVSL HGMDDPEREV LEWRFDSRLA FHHVARELHP EYFKNC
ARV 2     E       E      N      H      E  A  K      V     K       H          Y  D
LAV HAL   EE      E      NC     I  Q   E  A      K  K     S     LR  R    Q    Y  D
LAV ELI   QE    DTE      TN     ICQ    E     Q   K    N         E  K    N    FY  -
```

FIG. 3F

FIG. 4A

A

LAVbru vs.	GAG		POL		ENV Total		ENV OMP		ENV TMP	
HTLV-3 USA	512 0/0	0.8	1015 0/0	1.3	856 5/0	1.4	507 5/0	1.6	349 0/0	1.1
ARV-2 USA	502 12/2	3.4	1003 12/0	3.1	855 17/11	13.0	505 17/10	14.3	350 0/1	11.2
LAVeli Zaire	500 13/1	9.8	1002 13/0	5.5	853 22/14	20.7	504 22/14	25.3	349 0/0	13.8
LAVmal Zaire	505 14/7	12.0	1002 13/0	7.7	859 13/11	21.7	509 13/10	26.4	350 0/1	14.9

B

LAVeli vs.	GAG		POL		ENV Total		ENV OMP		ENV TMP	
LAVmal	505 1/6	10.8	1002 0/0	8.4	859 13/11	19.8	509 8/13	23.6	350 0/1	14.3

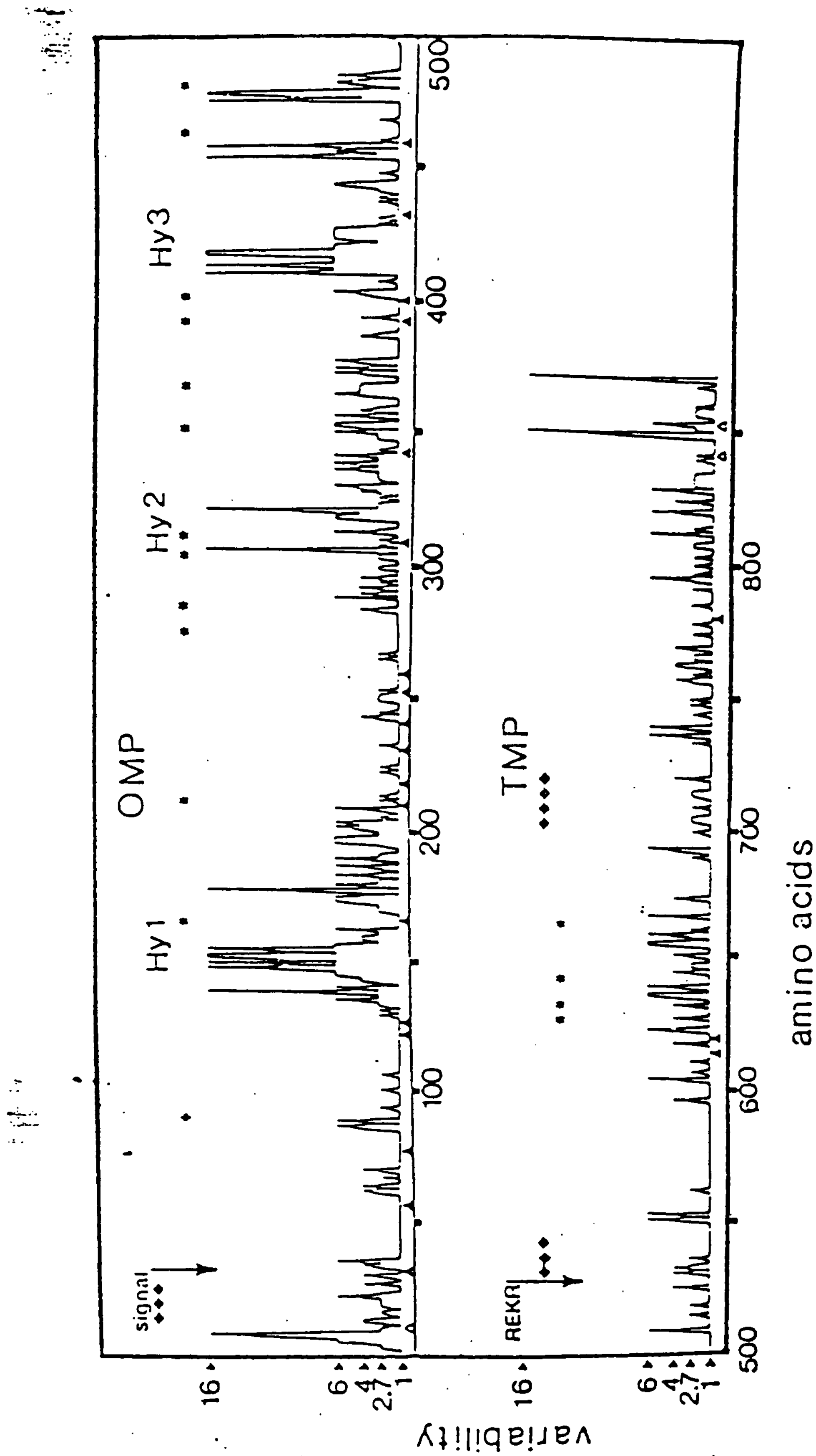
FIG. 4B

A

L'AVbru vs.	orf F		central region: orf Q		orf R		orf S	
HTLV-3 (USA)	206 0/0	1.5	192 0/0	0	,	nd	80 0/0	2.5
ARV-2 (USA)	210 0/4	12.6	192 0/0	10.0	97 0/1	9.4	81 0/1	15.0
LAVeli (Zaire)	206 1/1	19.4	192 0/0	10.4	96 0/0	11.5	80 0/0	27.5
LAVmal (Zaire)	209 2/5	27.0	192 0/0	12.6	96 0/0	10.4	80 0/0	23.8

B

LAVeli vs.	orf F		central region: orf Q		orf R		orf S	
LAVmal	209 3/6	22.5	192 0/0	12.0	96 0/0	6.3	80 0/0	11.3



OMP
TMP
Hy1
Hy2
Hy3
signal
REKR
variability
amino acids
16
6
4
2.7
1
100
200
300
400
500
600
700
800

FIG. 5

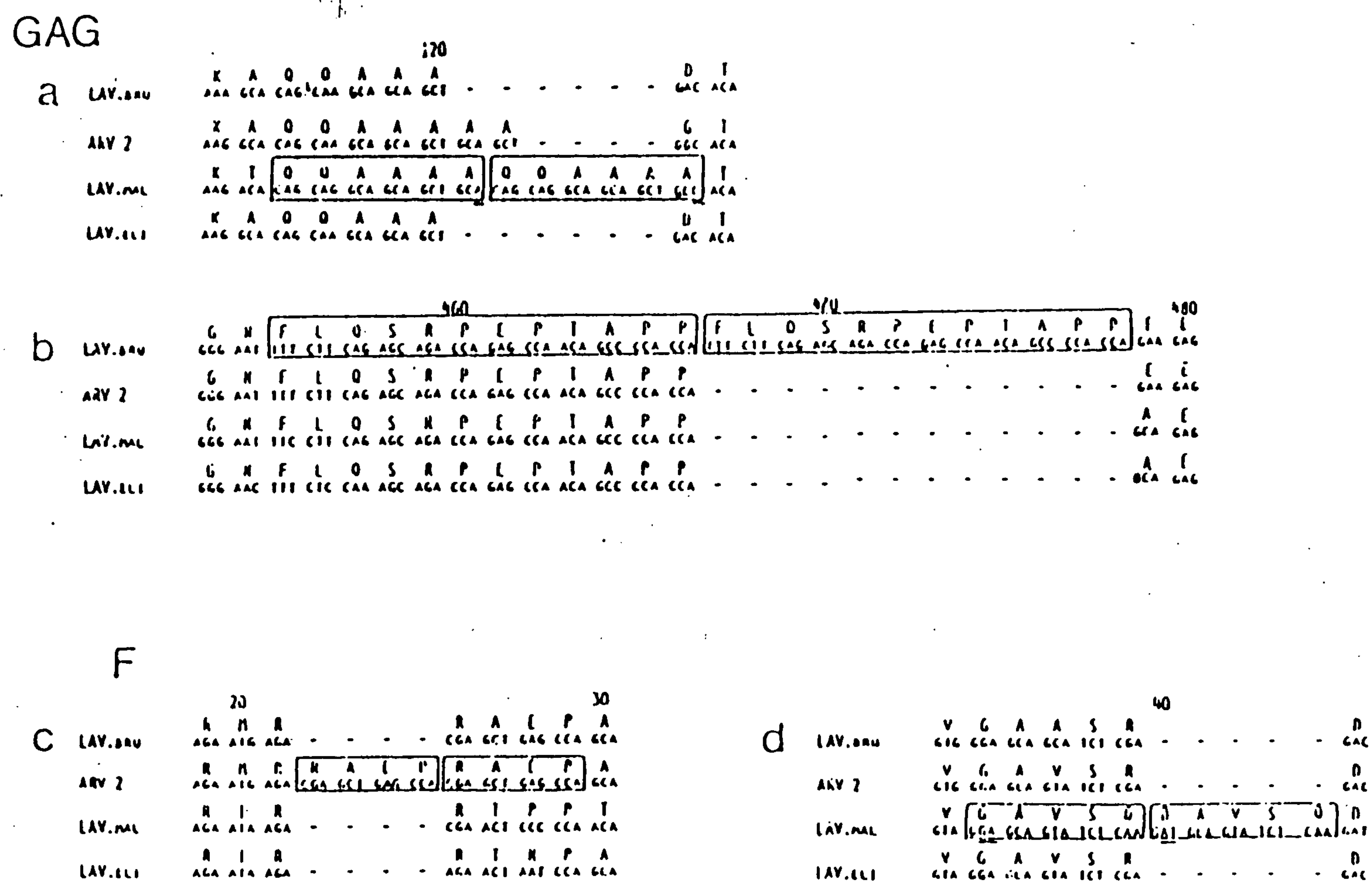


FIG. 6A

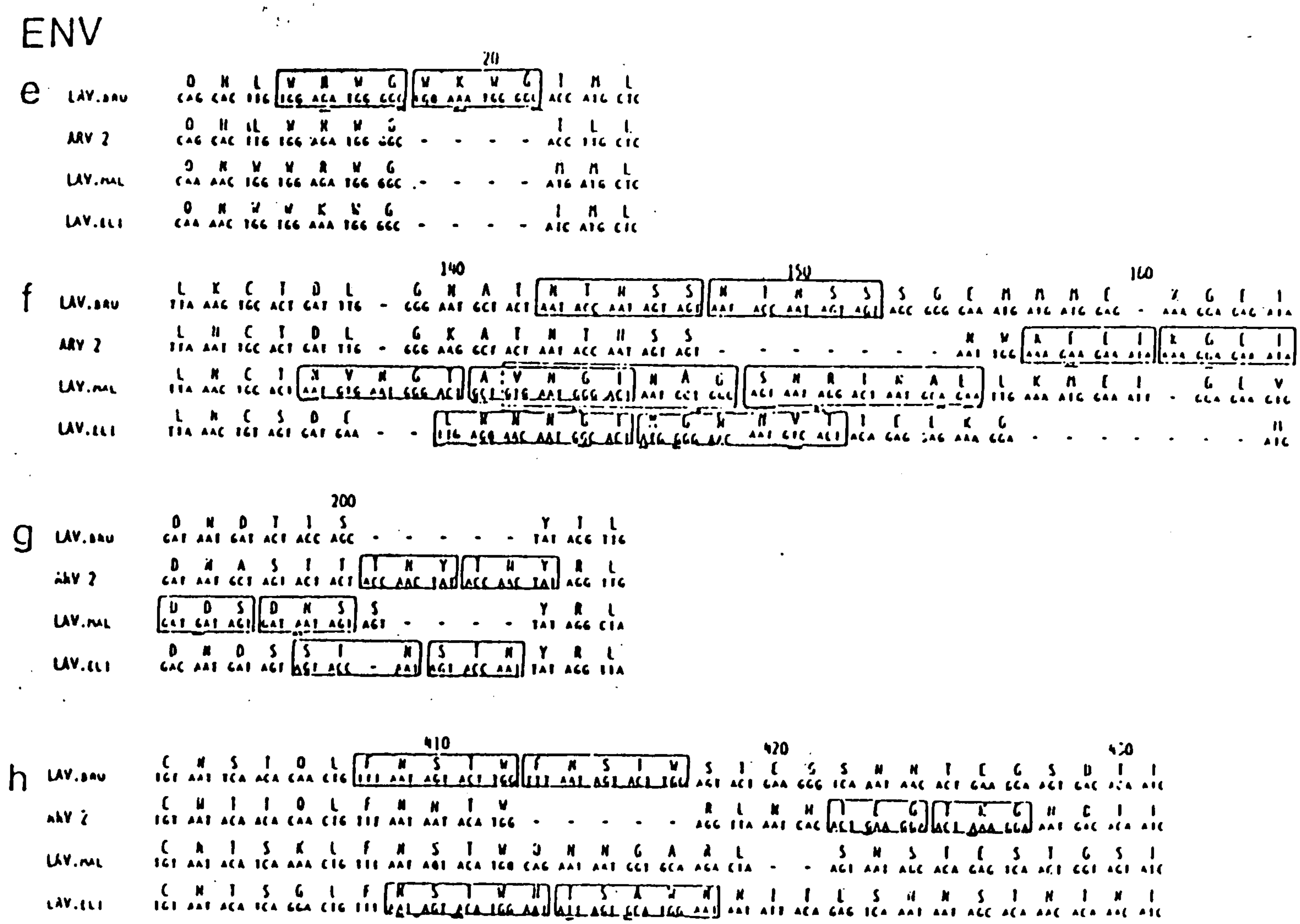

ENV
e
f
g
h
LAV.BRU
ARV 2
LAV.MAL
LAV.ELI
20
140
150
160
200
410
420
430

FIG. 6B

FIG. 7A

LAV.MAL

```
→R
GGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGCAAGGAACCCACTG
                                     R ↔ U5.
CTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAAGCAGTGTGTGCCCATCTGTTGTGT
                                      100              U5 ←
GACTCTGGTAACTAGAGATCCCTCAGACCACTCTAGACGGTGTAAAAATCTCTAGCAGTG

GCGCCCGAACAGGGACTTTAAAGTGAAAGTAACAGGGACTCGAAAGCGGAAGTTCCAGAG
             200
AAGTTCTCTCGACGCAGGACTCGGCTTGCTGAGGTGCACACAGCAAGAGGCGAGAGCGGC
                                                            300
                                                 →GAG
                                                 MetGlyAlaArg
GACTGGTGAGTACGCCAATTTTTGACTAGCGGAGGCTAGAAGGAGAGAGATGGGTGCGAG

AlaSerValLeuSerGlyGlyLysLeuAspAlaTrpGluLysIleArgLeuArgProGly
AGCGTCAGTATTAAGCGGGGGAAAATTAGATGCATGGGAGAAAATTCGGTTAAGGCCAGG
                                      400
GlyLysLysLysTyrArgLeuLysHisLeuValTrpAlaSerArgGluLeuGluArgPhe
GGGAAAGAAAAAATATAGACTGAAACATTTAGTATGGGCAAGCAGGGAGCTGGAAAGATT

AlaLeuAsnProGlyLeuLeuGluThrGlyGluGlyCysGlnGlnIleMetGluGlnLeu
CGCACTTAACCCTGGCCTTTTAGAAACAGGAGAAGGATGTCAACAAATAATGGAACAGCT
              500
GlnSerThrLeuLysThrGlySerGluGluIleLysSerLeuTyrAsnThrValAlaThr
ACAATCAACTCTCAAGACAGGATCAGAAGAAATTAAATCATTATATAATACAGTAGCAAC
                                                            600
LeuTyrCysValHisGlnArgIleAspValLysAspThrLysGluAlaLeuAspLysIle
CCTCTATTGTGTACATCAAAGGATAGATGTAAAAGACACCAAGGAAGCGCTAGATAAAAT

GluGluIleGlnAsnLysSerArgGlnLysThrGlnGlnAlaAlaAlaAlaGlnGlnAla
AGAGGAAATACAAAATAAGAGCAGGCAAAAGACACAGCAGGCAGCAGCTGCACAGCAGGC
                                      700
AlaAlaAlaThrLysAsnSerSerSerValSerGlnAsnTyrProIleValGlnAsnAla
AGCAGCTGCCACAAAAAACAGCAGCAGTGTCAGTCAAAATTACCCCATAGTGCAAAATGC

GlnGlyGlnMetIleHisGlnAlaIleSerProArgThrLeuAsnAlaTrpValLysVal
ACAAGGGCAAATGATACATCAGGCCATATCACCTAGGACTTTGAATGCATGGGTGAAAGT
              800
IleGluGluLysAlaPheSerProGluValIleProMetPheSerAlaLeuSerGluGly
AATAGAAGAAAAGGCTTTCAGCCCAGAAGTGATACCCATGTTCTCAGCATTATCAGAGGG
                                                            900
AlaThrProGlnAspLeuAsnMetMetLeuAsnIleValGlyGlyHisGlnAlaAlaMet
GGCCACCCCACAAGATTTAAATATGATGCTGAACATAGTTGGAGGACACCAGGCAGCTAT

GlnMetLeuLysAspThrIleAsnGluGluAlaAlaAspTrpAspArgValHisProVal
GCAAATGTTAAAAGATACCATCAATGAGGAAGCTGCAGACTGGGACAGGGTACATCCAGT
                                      1000
HisAlaGlyProIleProProGlyGlnMetArgGluProArgGlySerAspIleAlaGly
ACATGCAGGGCCTATTCCCCCAGGCCAGATGAGAGAACCAAGAGGAAGTGACATAGCAGG
```

FIG. 7B

ThrThrSerThrLeuGlnGluGlnIleGlyTrpMetThrSerAsnProProIleProVal
AACTACAAGTACCCTTCAAGAACAAATAGGATGGATGACAAGCAACCCACCTATCCCAGT
1100

GlyAspIleTyrLysArgTrpIleIleLeuGlyLeuAsnLysIleValArgMetTyrSer
GGGAGACATCTATAAAAGATGGATAATCCTGGGATTAAATAAAATAGTAAGAATGTATAG
1200

ProValSerIleLeuAspIleArgGlnGlyProLysGluProPheArgAspTyrValAsp
CCCTGTCAGCATTTTGGACATAAGACAAGGGCCAAAGGAACCTTTTAGAGACTATGTAGA

ArgPhePheLysThrLeuArgAlaGluGlnAlaThrGlnGluValLysAsnTrpMetThr
TAGGTTCTTTAAAACTCTCAGAGCTGAGCAAGCTACACAGGAGGTAAAAAATTGGATGAC
1300

GluThrLeuLeuValGlnAsnAlaAsnProAspCysLysThrIleLeuLysAlaLeuGly
AGAAACCTTGCTGGTCCAAAATGCGAATCCAGACTGTAAGACCATTTTAAAAGCATTAGG

ProGlyAlaThrLeuGluGluMetMetThrAlaCysGlnGlyValGlyGlyProSerHis
ACCAGGGGCTACATTAGAAGAAATGATGACAGCATGCCAGGGAGTGGGAGGACCCAGTCA
1400

LysAlaArgValLeuAlaGluAlaMetSerGlnAlaThrAsnSerThrAlaAlaIleMet
TAAAGCAAGAGTTTTGGCTGAGGCAATGAGCCAAGCAACAAATTCAACTGCTGCCATAAT
1500

MetGlnArgGlyAsnPheLysGlyGlnLysArgIleLysCysPheAsnCysGlyLysGlu
GATGCAGAGAGGTAATTTTAAGGGCCAGAAAAGAATTAAGTGTTTCAACTGTGGCAAAGA

GlyHisLeuAlaArgAsnCysArgAlaProArgLysLysGlyCysTrpLysCysGlyLys
AGGACACCTAGCCAGAAATTGCAGGGCCCCTAGGAAAAAGGGCTGTTGGAAATGTGGAA
1600 →POL

PhePheArgGluAsnLeu
GluGlyHisGlnMetLysAspCysThrGluArgGlnAlaAsnPheLeuGlyLysIleTrp
GGAAGGACACCAAATGAAAGACTGCACTGAGAGACAGGCTAATTTTTTAGGGAAAATTTG

AlaPheProGlnGlyLysAlaArgGluPheProSerGluGlnThrArgAlaAsnSerPro
ProSerHisLysGlyArgProGlyAsnPheLeuGlnSerArgProGluProThrAlaPro
GCCTTCCCACAAGGGAAGGCCAGGGAATTTCCTTCAGAGCAGACCAGAGCCAACAGCCCC
1700

ThrSerArgGluLeuArgValTrpGlyGlyAspLysThrLeuSerGluThrGlyAlaGlu
ProAlaGluSerPheGlyPheGlyGluGluIleLysProSerGlnLysGlnGluGlnLys
ACCAGCAGAGAGCTTCGGGTTTGGGGAGGAGATAAAACCCTCTCAGAAACAGGAGCAGAA
1800

ArgGlnGlyIleValSerPheSerPheProGlnIleThrLeuTrpGlnArgProValVal
AspLysGluLeuTyrProLeuAlaSerLeuLysSerLeuPheGlyAsnAspGlnLeuSer
AGACAAGGAATTGTATCCTTTAGCTTCCCTCAAATCACTCTTTGGCAACGACCAGTTGTC
GAG

ThrValArgValGlyGlyGlnLeuLysGluAlaLeuLeuAspThrGlyAlaAspAspThr
Gln
ACAGTAAGAGTAGGAGGACAGCTAAAAGAAGCTCTATTAGACACAGGAGCAGATGATACA
1900

ValLeuGluGluIleAsnLeuProGlyLysTrpLysProLysMetIleGlyGlyIleGly
GTATTAGAAGAAATAAATTTGCCAGGAAAATGGAAACCAAAAATGATAGGGGGAATTGGA

GlyPheIleLysValArgGlnTyrAspGlnIleLeuIleGluIleCysGlyLysLysAla
GGTTTTATCAAAGTAAGACAGTATGATCAAATACTTATAGAAATTTGTGGAAAAAAGGCT
2000

FIG. 7C

IleGlyThrIleLeuValGlyProThrProValAsnIleIleGlyArgAsnMetLeuThr
ATAGGTACAATATTGGTAGGACCTACACCTGTCAACATAATTGGACGAAATATGTTGACT
2100
GlnIleGlyCysThrLeuAsnPheProIleSerProIleGluThrValProValLysLeu
CAGATTGGTTGTACTTTAAATTTTCCAATTAGTCCTATTGAGACTGTACCAGTAAAATTA

LysProGlyMetAspGlyProArgValLysGlnTrpProLeuThrGluGluLysIleLys
AAGCCAGGGATGGATGGCCCAAGGGTTAAACAATGGCCATTGACAGAAGAAAAAATAAAA
2200
AlaLeuThrGluIleCysLysAspMetGluLysGluGlyLysIleLeuLysIleGlyPro
GCATTAACAGAAATTTGTAAAGATATGGAAAAGGAAGGAAAAATTTTAAAAATTGGGCCT

GluAsnProTyrAsnThrProValPheAlaIleLysLysLysAspSerThrLysTrpArg
GAAAATCCATACAATACTCCAGTATTTGCCATAAAGAAAAAAGACAGCACTAAATGGAGA
2300
LysLeuValAsnPheArgGluLeuAsnLysArgThrGlnAspPheTrpGluValGlnLeu
AAATTAGTGAATTTCAGAGAGCTTAATAAAAGAACTCAAGATTTTTGGGAAGTTCAATTA
2400
GlyIleProHisProAlaGlyLeuLysLysLysLysSerValThrValLeuAspValGly
GGAATACCACATCCTGCTGGGTTGAAAAAGAAAAAATCAGTCACAGTATTGGATGTGGGG

AspAlaTyrPheSerValProLeuAspGluAspPheArgLysTyrThrAlaPheThrIle
GATGCATATTTTTCAGTCCCTTTAGATGAAGATTTCAGGAAGTATACTGCATTCACTATA
2500
ProSerIleAsnAsnGluThrProGlyIleArgTyrGlnTyrAsnValLeuProGlnGly
CCCAGTATTAATAATGAGACACCAGGGATTAGATATCAGTACAATGTGCTACCACAGGGA

TrpLysGlySerProAlaIlePheGlnSerSerMetThrLysIleLeuGluProPheArg
TGGAAAGGATCACCAGCAATATTCCAGAGTAGCATGACAAAAATCTTAGAACCCTTTAGA
2600
ThrLysAsnProGluIleValIleTyrGlnTyrMetAspAspLeuTyrValGlySerAsp
ACAAAAAATCCAGAAATAGTCATATACCAATACATGGATGATTTGTATGTAGGGTCTGAT
2700
LeuGluIleGlyGlnHisArgThrLysIleGluGluLeuArgGluHisLeuLeuLysTrp
TTAGAAATAGGACAACATAGAACAAAAATAGAGGAACTAAGAGAACATCTATTGAAATGG

GlyPheThrThrProAspLysLysHisGlnLysGluProProPheLeuTrpMetGlyTyr
GGATTTACCACACCAGACAAAAAGCATCAGAAAGAACCCCCATTTCTTTGGATGGGGTAT
2800
GluLeuHisProAspLysTrpThrValGlnProIleGlnLeuProAspLysGluSerTrp
GAACTGCACCCTGACAAATGGACAGTGCAGCCTATACAACTGCCAGACAAGGAAAGCTGG

ThrValAsnAspIleGlnLysLeuValGlyLysLeuAsnTrpAlaSerGlnIleTyrPro
ACTGTCAATGATATACAGAAATTGGTGGGAAAACTAAATTGGGCAAGTCAGATTTATCCA
2900
GlyIleLysValLysGlnLeuCysLysLeuLeuArgGlyAlaLysAlaLeuThrAspIle
GGAATTAAAGTAAAGCAATTATGTAAACTCCTTAGGGGAGCAAAAGCACTAACAGACATA
3000
ValProLeuThrAlaGluAlaGluLeuGluLeuAlaGluAsnArgGluIleLeuLysGlu
GTACCATTAACTGCAGAGGCAGAATTAGAATTGGCAGAGAACAGGGAAATTCTAAAAGAA

FIG. 7D

ProValHisGlyValTyrTyrAspProSerLysAspLeuIleAlaGluIleGlnLysGln
CCAGTGCATGGGGTATATTATGACCCATCAAAAGACTTAATAGCAGAAATACAGAAGCAG
3100

GlyGlnGlyGlnTrpThrTyrGlnIleTyrGlnGluGlnTyrLysAsnLeuLysThrGly
GGGCAAGGTCAATGGACATATCAAATATACCAAGAGCAATATAAAAATCTGAAAACAGGG

LysTyrAlaArgIleLysSerAlaHisThrAsnAspValLysGlnLeuThrGluAlaVal
AAGTATGCAAGAATAAAGTCTGCCCACACTAATGATGTAAAACAATTAACAGAAGCAGTG
3200

GlnLysIleAlaGlnGluSerIleValIleTrpGlyLysThrProLysPheArgLeuPro
CAAAAGATAGCCCAAGAAAGCATAGTAATATGGGGAAAAACTCCTAAATTTAGACTACCC
3300

IleGlnLysGluThrTrpGluAlaTrpTrpThrGluTyrTrpGlnAlaThrTrpIlePro
ATACAAAAAGAAACATGGGAGGCATGGTGGACAGAATATTGGCAAGCCACCTGGATCCCT

GluTrpGluPheValAsnThrProProLeuValLysLeuTrpTyrGlnLeuGluThrGlu
GAATGGGAGTTTGTCAATACTCCTCCCCTAGTAAAACTATGGTACCAGTTAGAAACAGAA
3400

ProIleValGlyAlaGluThrPheTyrValAspGlyAlaAlaAsnArgGluThrLysLys
CCCATAGTAGGAGCAGAAACTTTCTATGTAGATGGGGCAGCTAATAGAGAAACTAAAAAG

GlyLysAlaGlyTyrValThrAspArgGlyArgGlnLysValValSerLeuThrGluThr
GGAAAAGCAGGATATGTTACTGACAGAGGAAGACAAAAGGTTGTCTCCTTAACTGAAACA
3500

ThrAsnGlnLysThrGluLeuGlnAlaIleHisLeuAlaLeuGlnAspSerGlySerGlu
ACAAATCAGAAGACTGAATTACAAGCAATCCACTTAGCTTTACAGGATTCAGGATCAGAA
3600

ValAsnIleValThrAspSerGlnTyrAlaLeuGlyIleIleGlnAlaGlnProAspLys
GTAAACATAGTAACAGACTCACAGTATGCATTAGGGATTATTCAAGCACAACCAGATAAA

SerGluSerGluIleValAsnGlnIleIleGluGlnLeuIleGlnLysAspLysValTyr
AGTGAATCAGAGATTGTTAATCAAATAATAGAGCAATTAATACAGAAGGACAAGGTCTAC
3700

LeuSerTrpValProAlaHisLysGlyIleGlyGlyAsnGluGlnValAspLysLeuVal
CTGTCATGGGTACCAGCACACAAAGGGATTGGAGGAAATGAACAAGTAGATAAATTAGTC

SerSerGlyIleArgLysValLeuPheLeuAspGlyIleAspLysAlaGlnGluGluHis
AGCAGTGGAATCAGAAAGGTACTATTTTTAGATGGGATAGATAAGGCTCAAGAAGAACAT
3800

GluLysTyrHisSerAsnTrpArgAlaMetAlaSerAspPheAsnLeuProProIleVal
GAAAAATATCACAGCAATTGGAGAGCAATGGCTAGTGACTTTAATCTACCACCTATAGTA
3900

AlaLysGluIleValAlaSerCysAspLysCysGlnLeuLysGlyGluAlaMetHisGly
GCGAAGGAAATAGTAGCCAGCTGTGATAAATGTCAACTAAAAGGGGAAGCCATGCATGGA

GlnValAspCysSerProGlyIleTrpGlnLeuAspCysThrHisLeuGluGlyLysIle
CAAGTAGACTGTAGTCCAGGGATATGGCAATTAGATTGCACACATCTAGAAGGAAAAATA
4000

IleIleValAlaValHisValAlaSerGlyTyrIleGluAlaGluValIleProAlaGlu
ATCATAGTAGCAGTCCATGTAGCCAGTGGATATATAGAAGCAGAAGTTATCCCAGCAGAA

ThrGlyGlnGluThrAlaTyrPheIleLeuLysLeuAlaGlyArgTrpProValLysVal
ACAGGACAGGAGACAGCATACTTTATACTAAAATTAGCAGGAAGATGGCCAGTAAAAGTA
4100

FIG. 7E

```
Val[illegible]rAspAsnGlySerAsnPheThrSerAlaAlaValLysAlaAlaCysTrpTrp
GTAC[illegible]AGACAATGGCAGCAATTTCACCAGTGCTGCAGTTAAAGCAGCCTGTTGGTGG
                                                            4200
AlaAsnIleLysGlnGluPheGlyIleProTyrAsnProGlnSerGlnGlyValValGlu
GCAAATATCAAACAGGAATTTGGAATTCCCTACAACCCCCAAAGTCAAGGAGTAGTGGAA

SerMetAsnLysGluLeuLysLysIleIleGlyGlnValArgGluGlnAlaGluHisLeu
TCTATGAATAAGGAATTAAAGAAAATCATAGGGCAGGTAAGAGAGCAAGCTGAACACCTT
                                          4300
LysThrAlaValGlnMetAlaValPheIleHisAsnPheLysArgLysGlyGlyIleGly
AAGACAGCAGTACAAATGGCAGTGTTCATTCACAATTTTAAAAGAAAAGGGGGGATTGGG

GlyTyrSerAlaGlyGluArgIleIleAspMetIleAlaThrAspIleGlnThrLysGlu
GGGTACAGTGCAGGGGAAAGAATAATAGACATGATAGCAACAGACATACAAACTAAAGAA
              4400
LeuGlnLysGlnIleThrLysIleGlnAsnPheArgValTyrTyrArgAspAsnArgAsp
TTACAAAAACAAATTACAAAAATTCAAAATTTTCGGGTTTATTACAGGGACAACAGAGAC
                                                            4500
ProIleTrpLysGlyProAlaLysLeuLeuTrpLysGlyGluGlyAlaValValIleGln
CCAATTTGGAAAGGACCAGCAAAACTACTCTGGAAAGGTGAAGGGGCAGTAGTAATACAG
                                                         Q-
AspAsnSerAspIleLysValValProArgArgLysAlaLysIleIleArgAspTyrGly
                                                         MetGlu
GACAATAGTGATATAAAGGTAGTACCAAGAAGAAAAGCAAAAATCATTAGGGATTATGGA
                                          4600   POL
LysGlnMetAlaGlyAspAspCysValAlaGlyGlyGlnAspGluAsp
 AsnArgTrpGlnValMetIleValTrpGlnValAspArgMetArgIleArgThrTrpHis
AAACAGATGGCAGGTGATGATTGTGTGGCAGGTGGACAGGATGAGGATTAGAACATGGCA

 SerLeuValLysHisHisMetTyrValSerLysLysAlaLysAsnTrpPheTyrArgHis
CAGTTTAGTAAAACATCATATGTATGTCTCAAAGAAAGCTAAAAATTGGTTTTATAGACA
              4700
 HisTyrGluSerArgHisProLysValSerSerGluValHisIleProLeuGlyAspAla
TCACTATGAAAGCAGGCATCCAAAAGTAAGTTCAGAAGTACACATCCCACTAGGGGATGC
                                                            4800
 ArgLeuValValArgThrTyrTrpGlyLeuGlnThrGlyGluLysAspTrpHisLeuGly
TAGATTAGTAGTAAGAACATATTGGGGTCTGCAAACAGGAGAAAAAGACTGGCACTTGGG

 HisGlyValSerIleGluTrpArgGlnLysArgTyrSerThrGlnLeuAspProAspLeu
TCATGCGGTCTCCATAGAATGGAGGCAGAAAAGATATAGCACACAACTAGATCCTGACCT
                                          4900
 Al[illegible]pGlnLeuIleHisLeuTyrTyrPheAspCysPheSerGluSerAlaIleArgGln
AGCAGACCAACTGATTCATCTGTACTATTTTGATTGTTTTTCAGAATCTGCCATAAGACA

 AlaIleLeuGlyHisIleValSerProArgCysAspTyrGlnAlaGlyHisAsnLysVal
AGCCATATTAGGACATATAGTTAGTCCTAGGTGTGATTATCAAGCAGGACATAACAAGGT
              5000
 GlySerLeuGlnTyrLeuAlaLeuThrAlaLeuIleAlaProLysLysThrArgProPro
AGGATCTTTACAGTATTTGGCACTAACAGCATTAATAGCACCAAAAAAGACAAGGCCACC
                                        R                   5100
                                        MetGluGlnAlaProAlaAspGlnGly
 LeuProSerValArgLysLeuThrGluAspArgTrpAsnLysProGlnGlnThrLysGly
TTTGCCTAGTGTTAGGAAGCTAACAGAAGATAGATGGAACAAGCCCCAGCAGACCAAGGG
```

FIG. 7F

```
                                            Q
ProGlnArgGluProHisAsnGluTrpThrLeuGluLeuLeuGluGluLeuLysGlnGlu
 HisArgGlyS rHisThrMetAsnGlyHis
CCACAGAGGGAGCCACACAATGAATGGACATTAGAACTTTTAGAGGAGCTTAAGCAAGAA
                                      5200

AlaValArgHisPheProArgIleTrpLeuHisSerLeuGlyGlnHisIleTyrGluThr
GCTGTCAGACACTTTCCTAGGATATGGCTCCATAGTTTAGGACAACATATCTATGAAACT

TyrGlyAspThrTrpGluGlyValGluAlaIleIleArgSerLeuGlnGlnLeuLeuPhe
TATGGGGATACCTGGGAAGGAGTTGAAGCTATAATAAGAAGTCTGCAACAACTGCTGTTT
            5300

IleHisPheArgIleGlyCysGlnHisSerArgIleGlyIleThrArgGlnArgArgAla
ATTCATTTCAGAATTGGGTGTCAACATAGCAGAATAGGCATTACTCGACAGAGAAGAGCA
      S           R                                     5400
ArgAsnGlySerSerArgSer
    MetAspProValAspProAsnLeuGluProTrpAsnHisProGlySerGlnProArg
AGAAATGGATCCAGTAGATCCTAACTTAGAGCCCTGGAACCATCCAGGGAGTCAGCCTAG

 ThrProCysAsnLysCysTyrCysLysLysCysCysTyrHisCysGlnMetCysPheIle
GACGCCTTGTAATAAGTGTTATTGTAAAAAGTGCTGCTATCATTGCCAAATGTGCTTCAT
                                      5500

 ThrLysGlyLeuGlyIleSerTyrGlyArgLysLysArgArgGlnArgArgArgProPro
AACGAAAGGCTTAGGCATCTCCTATGGCAGGAAGAAGCGGAGACAGCGACGAAGACCTCC
                                      S
 GlnGlyAsnGlnAlaHisGlnAspProLeuProGluGln
TCAGGGCAATCAGGCTCATCAAGATCCTCTACCAGAGCAGTAAGTAGTATATGTAATACA
            5600

ACCTTTAGTGATATTAGCAATAGTAGCATTAGTAGTAACGCTAATAATAGCAATAGTTGT
                                                        5700

GTGGACCATAGTATTTATAGAAATTAGGAAAATAAGAAGACAAAGGAAAATAGACAGGTT

                                        ENV
                                        MetArgValArgGluIleGlnArg
GATTGATAGAATAAGAGAAAGCAGAAGATAGTGGCAATGAGAGTGAGGGAGATACAGA
                                      5800

  AsnTyrGlnAsnTrpTrpArgTrpGlyMetMetLeuLeuGlyMetLeuMetThrCysSer
GGAATTATCAAAACTGGTGGAGATGGGGCATGATGCTCCTTGGGATGTTGATGACCTGTA

  IleAlaGluAspLeuTrpValThrValTyrTyrGlyValProValTrpLysGluAlaThr
GTATTGCAGAAGATTTGTGGGTTACAGTTTATTATGGGGTACCTGTGTGGAAAGAAGCAA
            5900

  ThrThrLeuPheCysAlaSerAspAlaLysSerTyrGluThrGluValHisAsnIleTrp
CCACTACCCTATTTTGTGCATCAGATGCTAAATCATATGAAACAGAAGTACATAACATCT
                                                        6000

  AlaThrHisAlaCysValProThrAspProAsnProGlnGluIleGluLeuGluAsnVal
GGGCTACACATGCCTGTGTACCCACGGACCCCAACCCACAAGAAATAGAACTGGAAAATG

  ThrGluGlyPheAsnMetTrpLysAsnAsnMetValGluGlnMetHisGluAspIleIle
TCACAGAAGGGTTTAACATGTGGAAAAATAACATGGTGGAGCAGATGCATGAGGATATAA
                                      6100
```

FIG. 7G

```
SerLeuTrpAspGlnSerLeuLysPr CysValLysLeuThrProLeuCysValThrLeu
TCAGTTTATGGGATCAAAGCCTAAAACCATGTGTAAAGCTAACCCCACTCTGTGTCACTT

AsnCysThrAsnValAsnGlyThrAlaValAsnGlyThrAsnAlaGlySerAsnArgThr
TAAACTGCACTAATGTGAATGGGACTGCTGTGAATGGGACTAATGCTGGGAGTAATAGGA
            6200
AsnAlaGluLeuLysMetGluIleGlyGluValLysAsnCysSerPheAsnIleThrPro
CTAATGCAGAATTGAAAATGGAAATTGGAGAAGTGAAAAACTGCTCTTTCAATATAACCC
                                                        6300
ValGlySerAspLysArgGlnGluTyrAlaThrPheTyrAsnLeuAspLeuValGlnIle
CAGTAGGAAGTGATAAAAGGCAAGAATATGCAACTTTTTATAACCTTGATCTAGTACAAA

AspAspSerAspAsnSerSerTyrArgLeuIleAsnCysAsnThrSerValIleThrGln
TAGATGATAGTGATAATAGTAGTTATAGGCTAATAAATTGTAATACCTCAGTAATTACAC
                                    6400
AlaCysProLysValThrPheAspProIleProIleHisTyrCysAlaProAlaGlyPhe
AGGCTTGTCCAAAGGTAACCTTTGATCCAATTCCCATACATTATTGTGCCCCAGCTGGTT

AlaIleLeuLysCysAsnAspLysLysPheAsnGlyThrGluIleCysLysAsnValSer
TTGCAATTCTAAAGTGTAATGATAAGAAGTTCAATGGAACGGAAATATGTAAAAATGTCA
            6500
ThrValGlnCysThrHisGlyIleLysProValValSerThrGlnLeuLeuLeuAsnGly
GTACAGTACAATGTACACATGGAATTAAGCCAGTGGTGTCAACTCAACTGCTGTTAAATG
                                                        6600
SerLeuAlaGluGluGluIleMetIleArgSerGluAsnLeuThrAspAsnThrLysAsn
GCAGTCTAGCAGAAGAAGAGATAATGATTAGATCTGAAAATCTCACAGACAATACTAAAA

IleIleValGlnLeuAsnGluThrValThrIleAsnCysThrArgProGlyAsnAsnThr
ACATAATAGTACAGCTTAATGAAACTGTAACAATTAATTGTACAAGGCCTGGAAACAATA
                                    6700
ArgArgGlyIleHisPheGlyProGlyGlnAlaLeuTyrThrThrGlyIleValGlyAsp
CAAGAAGAGGGATACATTTCGGCCCAGGGCAAGCACTCTATACAACAGGGATAGTAGGAG

IleArgArgAlaTyrCysThrIleAsnGluThrGluTrpAspLysThrLeuGlnGlnVal
ATATAAGAAGAGCATATTGTACTATTAATGAAACAGAATGGGATAAAACTTTACAACAGG
            6800
AlaValLysLeuGlySerLeuLeuAsnLysThrLysIleIlePheAsnSerSerSerGly
TAGCTGTAAAACTAGGAAGCCTTCTTAACAAAACAAAAATAATTTTTAATTCATCCTCAG
                                                        6900
GlyAspProGluIleThrThrHisSerPheAsnCysArgGlyGluPhePheTyrCysAsn
GAGGGGACCCAGAAATTACAACACACAGTTTTAATTGTAGAGGGGAATTTTTCTACTGTA

ThrSerLysLeuPheAsnSerThrTrpGlnAsnAsnGlyAlaArgLeuSerAsnSerThr
ATACATCAAAACTGTTTAATAGTACATGGCAGAATAATGGTGCAAGACTAAGTAATAGCA
                                    7000
GluSerThrGlySerIleThrLeuProCysArgIleLysGlnIleIleAsnMetTrpGln
CAGAGTCAACTGGTAGTATCACACTCCCATGCAGAATAAAACAAATTATAAATATGTGGC

LysThrGlyLysAlaMetTyrAlaProProIleAlaGlyValIleAsnCysLeuSerAsn
AGAAAACAGGAAAAGCTATGTATGCCCCTCCCATCGCAGGAGTCATCAACTGTTTATCAA
            7100
IleThrGlyLeuIleLeuThrArgAspGlyGlyAsnSerSerAspAsnSerAspAsnGlu
ATATTACAGGGCTGATATTAACAAGAGATGGTGGAAATAGTAGTGACAATAGTGACAATG
                                                        7200
```

FIG. 7H

```
 ThrLeuArgPr GlyGlyGlyAspMetArgAspAsnTrpIleSerGluLeuTyrLysTyr
AGACCTTAAGACCTGGAGGAGGAGATATGAGGGACAATTGGATAAGTGAATTATATAAAT

 LysValValArgIleGluProLeuGlyValAlaProThrLysAlaLysArgArgValVal
ATAAAGTAGTAAGAATTGAACCCCTAGGAGTAGCACCCACCAAGGCAAAGAGAAGAGTGG
                                      7300
 GluArgGluLysArgAlaIleGlyLeuGlyAlaMetPheLeuGlyPheLeuGlyAlaAla
TGGAAAGAGAAAAAAGAGCAATAGGACTAGGAGCCATGTTCCTTGGGTTCTTGGGAGCAG

 GlySerThrMetGlyAlaAlaSerLeuThrLeuThrValGlnAlaArgGlnLeuLeuSer
CAGGAAGCACGATGGGCGCAGCGTCACTAACGCTGACGGTACAGGCCAGACAGTTACTGT
              7400
 GlyIleValGlnGlnGlnAsnAsnLeuLeuArgAlaIleGluAlaGlnGlnHisLeuLeu
CTGGTATAGTGCAACAGCAAAACAATTTGCTGAGGGCTATAGAGGCGCAACAGCATCTGT
                                                        7500
 GlnLeuThrValTrpGlyIleLysGlnLeuGlnAlaArgValLeuAlaValGluArgTyr
TGCAACTCACGGTCTGGGGCATTAAACAGCTCCAGGCAAGAGTCCTGGCTGTGGAAAGAT

 LeuGlnAspGlnArgLeuLeuGlyMetTrpGlyCysSerGlyLysHisIleCysThrThr
ACCTACAGGATCAACGGCTCCTAGGAATGTGGGGTTGCTCTGGAAAACACATTTGCACCA
                                      7600
 PheValProTrpAsnSerSerTrpSerAsnArgSerLeuAspAspIleTrpAsnAsnMet
CATTTGTGCCTTGGAACTCTAGTTGGAGTAATAGATCTCTAGATGACATTTGGAATAATA

 ThrTrpMetGlnTrpGluLysGluIleSerAsnTyrThrGlyIleIleTyrAsnLeuIle
TGACCTGGATGCAGTGGGAAAAAGAAATTAGCAATTACACAGGCATAATATACAACTTAA
              7700
 GluGluSerGlnIleGlnGlnGluLysAsnGluLysGluLeuLeuGluLeuAspLysTrp
TTGAAGAATCGCAAATCCAGCAAGAAAAGAATGAAAAGGAATTATTGGAATTGGACAAGT
                                                        7800
 AlaSerLeuTrpAsnTrpPheSerIleSerLysTrpLeuTrpTyrIleArgIlePheIle
GGGCAAGTTTGTGGAATTGGTTTAGCATATCAAAATGGCTGTGGTATATAAGAATATTCA

 IleValValGlyGlyLeuIleGlyLeuArgIleIlePheAlaValLeuSerLeuValAsn
TAATAGTAGTAGGAGGCTTAATAGGTTTAAGAATAATTTTTGCTGTGCTTTCTTTAGTAA
                                      7900
 ArgValArgGlnGlyTyrSerProLeuSerLeuGlnThrLeuLeuProThrProArgGly
ATAGAGTTAGGCAGGGATACTCACCTCTGTCGTTGCAGACCCTCCTCCCAACACCGAGGG

 ProProAspArgProGluGlyIleGluGluGlyGlyGluGlnGlyArgGlyArgSer
GACCACCCCGACAGGCCCGAAGGAATAGAAGAAGAAGGTGGAGAGCAAGGCAGAGGCAGAT
              8000
 IleArgLeuValAsnGlyPheSerAlaLeuIleTrpAspAspLeuArgAsnLeuCysLeu
CAATTCGATTGGTGAACGGATTCTCAGCACTTATCTGGGACGACCTGAGGAACCTGTGCC
                                                        8100
 PheSerTyrHisArgLeuArgAspLeuLeuLeuIleAlaThrArgIleValGluLeuLeu
TCTTCAGTTACCACCGCTTGAGAGACTTACTCTTAATTGCAACGAGGATTGTGGAACTTC

 GlyArgArgGlyTrpGluAlaLeuLysTyrLeuTrpAsnLeuLeuGlnTyrTrpGlyGln
TGGGACGCAGGGGGTGGGAAGCCCTCAAATATCTGTGGAATCTCCTGCAATATTGGGGTC
                                      8200
```

FIG. 7I

GluLeuLysAsnSerAlaIleSerLeuLeuAsnThrThrAlaIleAlaValAlaGluCys
AGGAACTGAAGAATAGTGCTATTAGCTTGCTTAATACCACAGCAATAGCAGTAGCTGAAT

ThrAspArgValIleGluIleGlyGlnArgPheGlyArgAlaIleLeuHisIleProArg
GCACAGATAGGGTTATAGAAATAGGACAAAGATTTGGTAGAGCTATTCTCCACATACCTA
8300

F
MetGlyGlyLysTrpSerLys
EW
ArgIleArgGlnGlyPheGluArgAlaLeuLeu
GAAGAATTAGACAGGGCTTCGAAAGGGCTTTGCTATAACATGGGTGGCAAGTGGTCAAAA
8400

SerSerIleValGlyTrpProLysIleArgGluArgIleArgArgThrProProThrGlu
AGTAGCATAGTAGGATGGCCTAAGATTAGGGAAAGAATAAGACGAACTCCCCCAACAGAA

ThrGlyValGlyAlaValSerGlnAspAlaValSerGlnAspLeuAspLysCysGlyAla
ACAGGAGTAGGAGCAGTATCTCAAGATGCAGTATCTCAAGATTTAGATAAATGTGGAGCA
8500

AlaAlaSerSerSerProAlaAlaAsnAsnAlaSerCysGluProProGluGluGluGlu
GCCGCAAGCAGCAGTCCAGCAGCTAATAATGCTAGTTGTGAACCACCAGAAGAAGAGGAG

GluValGlyPheProValArgProGlnValProLeuArgProMetThrTyrLysGlyAla
GAGGTAGGCTTTCCAGTCCGTCCTCAGGTACCTTTAAGACCAATGACTTATAAAGGAGCT
8600

U3
PheAspLeuSerHisPheLeuLysGluLysGlyGlyLeuAspGlyLeuValTrpSerPro
TTTGATCTCAGCCACTTTTTAAAAGAAAAGGGGGGACTGGATGGGTTAGTTTGGTCCCCA
8700

LysArgGlnGluIleLeuAspLeuTrpValTyrHisThrGlnGlyTyrPheProAspTrp
AAAAGACAAGAAATCCTTGATCTGTGGGTCTACCACACACAAGGCTACTTCCCTGATTGG

GlnAsnTyrThrProGlyProGlyIleArgPheProLeuThrPheGlyTrpCysPheLys
CAGAATTACACACCAGGGCCAGGGATTAGATTCCCACTGACCTTCGGATGGTGCTTTAAG
8800

LeuValProMetSerProGluGluValGluGluAlaAsnGluGlyGluAsnAsnCysLeu
TTAGTACCAATGAGTCCAGAGGAAGTAGAGGAGGCCAATGAAGGAGAGAACAACTGTCTG

LeuHisProIleSerGlnHisGlyMetGluAspAlaGluArgGluValLeuLysTrpLys
TTACACCCTATTAGCCAACATGGAATGGAGGACGCAGAAAGAGAAGTGCTAAAATGGAAG
8900

PheAspSerSerLeuAlaLeuArgHisArgAlaArgGluGlnHisProGluTyrTyrLys
TTTGACAGCAGCCTAGCACTAAGACACAGAGCCAGAGAACAACATCCGGAGTACTACAAA
9000

F
AspCys
GACTGCTGACACAGAAGTTGCTGACAGGGGACTTTCCGCTGGGGACTTTCCAGGGGAGGC

GTAACTTGGGCGGGACCGGGGAGTGGCTAACCCTCAGATGCTGCATATAAGCAGCTGCTT
9100

U3 R
TTCGCCTGTACTGGGTCTCTCTTGTTAGACCAGGTCGAGCCCGGGAGCTCTCTGGCTAGC

AAGGAACCCACTGCTTAAGCCTCAATAAAGCTTGCCTTGAGTGCCTCAA
9200